GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2000, 20:43:04 ; Search time 284.47 Seconds
(without alignments)
3724.693 Million cell updates/sec

Title: US-09-174-937-4
Perfect score: 4235
Sequence: 1 gtcgacccacgcgtccgcgg.....aaaaaaaaaagggcggccgc 4235

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
C 1	692.8	16.4	783	1	X39880	Gastric cancer ass
2	332	7.8	1227	1	X39873	Gastric cancer ass
3	332	7.8	816	1	X39876	Gastric cancer ass
4	298	7.0	300	1	T23651	Human gene signatu
5	295.2	6.9	1386	1	T39229	Apoptosis-regulati
6	286.2	6.8	1386	1	X04120	Human Nip2 encodin
C 7	63.8	1.5	5952	1	Q11710	Dictyostelium plas
C 8	62	1.5	1536	1	V44964	Clone CQ331_2 codi
9	61.6	1.5	2280	1	V52604	Human metalloprote
10	60	1.4	4237	1	V61487	Human secreted pro
11	59.8	1.4	2447	1	V54597	Human secretory pr
12	59.6	1.4	1474	1	T90174	Oil seed rape cyst
13	59	1.4	1834	1	Q04690	Encodes Mammalian
14	58.8	1.4	1172	1	Q58669	Protein allergen o
15	58.6	1.4	240	1	T76782	Staphylococcus aur
16	58.4	1.4	4358	1	V42653	Nucleotide sequenc
17	58.4	1.4	2634	1	X27062	S. tuberosum isoam
C 18	57.2	1.4	5760	1	N50530	Sequence encoding
19	57.2	1.4	1582	1	T18831	Human survival mot
20	57.2	1.4	1560	1	T18828	Human survival mot
21	57.2	1.4	1582	1	T28259	Survival motor neu
22	57.2	1.4	1733	1	V81394	Human tumour antig
23	57	1.3	1558	1	Q10896	Encodes Xenopus Bo
24	57	1.3	1461	1	T01471	Mammalian mast cel
25	56.8	1.3	2123	1	T59701	PTH-like peptide D
26	56.8	1.3	2205	1	V04680	Human presenilin i
C 27	56.8	1.3	1875	1	V33033	Dirofilaria immiti
28	56.8	1.3	1875	1	V33017	Dirofilaria immiti
29	56.8	1.3	1549	1	X24921	Guman ras carboxy-
30	56.8	1.3	2120	1	X33812	Coding sequence fo
31	56.6	1.3	1641	1	T15758	Human interleukin-
32	56.4	1.3	2427	1	Q04107	Human pro-urokinas
C 33	56.2	1.3	3101	1	Q02047	Sequence encoding
34	56.2	1.3	1558	1	T28255	Survival motor neu
35	55.8	1.3	3684	1	X21560	Human N-arginine d
36	55.6	1.3	3905	1	N70461	Sequence encoding
37	55.6	1.3	3905	1	Q25893	Human Factor XIII
38	55.6	1.3	2682	1	V63195	cDNA from clone fg
39	55.4	1.3	1410	1	Q15022	Hyoscyamine 6 beta
40	55.4	1.3	3410	1	V58586	Prostate tumour sp
41	55.4	1.3	618	1	V58362	Polio: sequence fo
42	55.4	1.3	3410	1	V61201	Full length cDNA s
43	55.2	1.3	12827	1	V19136	Equine arteritis v
44	55.2	1.3	15528	1	V19059	Equine arteritis v
45	55	1.3	2549	1	X91467	Sequence of human

ALIGNMENTS

RESULT 1
X39880/c
ID X39880 standard; DNA; 783 BP.
AC X39880;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides f
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67; Page 582; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 783 BP; 259 A; 148 C; 190 G; 249 T;

Query Match 16.4%; Score 692.8; DB 1; Length 783;
Best Local Similarity 98.2%; Pred. No. 1.9e-191;
Matches 721; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

Qy 3455 gtgctagggaaaaagctaccatgtacatt...ttacc.caaatac.tagcatcttaagatgaa 3511
Db 734 GGGCTAAGGAAAAGCTACCATGTTCATT...ATCCCTAACCTTAATGCATCCTACATCAA 675

Qy 3512 tcc-tagaaaaatataatcccttaa...ttaa...ttttgtttttta 3570
Db 674 TCCNTAGAAAAATAATCCNTAAGCAGATAGGTAGNCAGATCTAACCATTCATATGATTTA 615

Qy 3571 gctctctagctcttgcactctgaacattcttgcttt...ttctcaggaactgctt 3630
Db 614 GCTCTCTAGCTCTTGCACTCTGAACATTCTTGCTTTGGTTCTGATTCTTCGCAACTGCTT 555

Qy 3631 tgcatttctcctatagatctgtag...ggcctcaggggtcttagggcaaaagcatt 3690

```
Db    554 TGCATTTCTCCTATAGATCTGTAGTTAATGGAACCAAGGGGTCATTGGGGCAAAAGCATT 495

Qy   3691 gtttctcaaagctcctttgattaaagagaaaaaaagaaaatttacacagaagatagtgtcaa 3750
          ||| |||||||| | ||||||||| || |||||||||||||||||||||||||||||||||
Db    494 GTTTCTCAAAGCTCCTTGATTAAAGAGAAAGAACAGAAATTTGCACAGAAGATAGTGTCAA 435

Qy   3751 ggagtgagaaagtttgtttgagggcagtagcttcagtgtggaagaaaatcctgaagtttct 3810
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    434 GGAGTGAGAAAGTTTGTTTGAGGGCAGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCT 375

Qy   3811 gttgaaggcatacaatgttctatggggttactctctaagacattctctgaggtgtgtgag 3870
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    374 GTTGAAGGCATACAATGTTCTATGGGGTTACTCTCTAAGACATTCTCTGAGGTGTGTGAG 315

Qy   3871 gaagtcactactcctagcctttgttaagatgtaatttttaaatattcagttatggtactat 3930
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    314 GAAGTCACTACTCCTAGCCTTTGTTAAGATGTAATTTTTAAATATTCAGTTATGGTACTAT 255

Qy   3931 gtttgcaagctctcgtcttatcacaatgcctcagtagtttgttcccttagaaacatttaga 3990
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    254 GTTTGCAAGCTCTCGTCTTATCACAATGCCTCAGTAGTTTGTTCCCTTAGAAACATTTAGA 195

Qy   3991 tgtgcacaaattaatcttttatatatctaaaggtttttctatcatgcattggattgctca 4050
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    194 TGTGCACAAATTAATCTTTTATATATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCA 135

Qy   4051 gaataaaagtgtctgttagacttcgttttggtaaaataaattctccataatgtagattaata 4110
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    134 GAATAAAGTGTCTGTTAGACTTCGTTTTGGTAAATAAATTCTCCATAATGTAGATTAATA 75

Qy   4111 atataaaagtctttaatgacacaatatatctatatagcctcactgtataattcagaaata 4170
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db     74 ATATAAAAGTCTTTAATGACACAATATATCTATATAGCCTCACTGTATAATTCAGAAATA 15

Qy   4171 aaaattgattctgc 4184
          ||||||||||||||
Db     14 AAAATTGATTCTGC 1


RESULT   2
X39873
ID   X39873 standard; DNA; 1227 BP.
AC   X39873;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 579-580; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises. (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   1227 BP;    388 A;    270 C;    304 G;    264 T;


  Query Match             7.8%;  Score 332;  DB 1;  Length 1227;
  Best Local Similarity   94.3%;  Pred. No. 1.6e-58;
  Matches  378;  Conservative    0;  Mismatches   16;  Indels    7;  Gaps    3;

Qy     16 cgcggaaatggaggaggagacagaagttccttgagctcggaaccaggatatcaaaaaaaa 75
          | |||||||||||||||||||||||||| | ||| ||||||||| ||||||||||||||
Db    818 CACGGAAATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGAACCAGGATATCAAAAAAAA 877

Qy     76 tggactactgtcagaggatgtaggaatggacatcccctttgaaagagggcgtgctgagtct 135
          |||||||||||||||||||||||||| | |||| |||||| || |||| | | | | |||
Db    878 TGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCT 937

Qy    136 cagtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactgtatc 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    938 CAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTGTATC 997

Qy    196 tcggagaatcaagctcacagccccaaatatcaatctttctctggaccaaaagtgaaagaat 255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    998 TCGGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAAGTGAAAGAT 1057

Qy    256 tattctctctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaaa 314
          |||||||||||||||||| ||||| ||||||||||||| |||||||||||||||||| |
Db   1058 TATTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAAT 1117

Qy    315 ttgatacccccgatgaagcagattcttttgagtacactgccc----atgatcccaggagcc 370
          |||||||||||||||||||||||||| |||||||| |||||    ||||||||||||| |
Db   1118 TTGATACCCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCCCATGAAGAATCCCAGAGT 1177

Qy    371 aacaaagattctggcc--aagagtcagagctctattccagaa 409
          |||||||||||||||  | ||||||||||||||| | | |
Db   1178 CACAAAGATTCTGGCCCAAGAAGTCAGAGTCTATTTCNGGA 1218


RESULT   3
X39876
ID   X39876 standard  DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581; 787pp; English.
```

```
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    209 G;    184 T;

  Query Match             7.8%;  Score 332;  DB 1;  Length 816;
  Best Local Similarity   94.3%;  Pred. No. 1.4e-58;
  Matches  378;  Conservative    0;  Mismatches   16;  Indels    7;  Gaps    3;

Qy      16 cgcggaaatggaggaggagacagagttccttgagctcggaaccaggatatcaagaccaaa 75
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407 CAGGAAAIGGAGGAGGAGACAGAGTTCCTTCAGCTCGGAACCAGGATATCAAGACCAAA 466

Qy      76 tggactactgtcagaggatgtaggaatggacatccccttttgaagagggcgtgctgagtcc 135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     467 TGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCC 526

Qy     136 cagtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcc 195
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 CAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCC 586

Qy     196 tcggagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatc 255
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     587 TCGGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATC 646

Qy     256 tattctctctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaac 314
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     647 TATTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAAC 706

Qy     315 ttgataccccccgatgaagcagattcttttgagtcacactggcc----atgatcccacagcc 370
           |||||||||||||||||||||||||||||||||||||||||||||    |||||||||||||
Db     707 TTGATACCCCCGATGAAGCAGATTCTTTTGACTACCCTGGCCCATGAAGAATCCCACAGC 766

Qy     371 aacaaagattctggcc--aagagtcagagtctattccagaa 409
           |||||||||||||||  |  ||||||| |||| | | |
Db     767 CACAAAGATTCTGGCCCAAGAAGTCAGAGTCTATTTCNGGA 807


RESULT   4
T23651
ID   T23651 standard; cDNA to mRNA; 300 BP.
AC   T23651;
DT   17-SEP-1995  (first entry)
DE   Human gene signature HUMGS05515.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; 301916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e g
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1421; 2245pp; Japanese.

CC   A single-stranded DNA (or its complementary strand or RNA thereof),
CC   double-stranded DNA) which comprises one of the "5'" "3'" sequences
CC   given in T19001-T26937 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
CC   sequences were obtained from 3' directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(A) as the sole primer. Since the 5'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   300 BP;    89 A;    44 C;    74 G;    92 T;

  Query Match             7.0%;  Score 298;  DB 1;  Length 816;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-52;
  Matches  298;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    3646 gatctgtagttaagggaaccaagggtcattgggaaaaaattttttaaaaaatt 3705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GATCTGTAGTTAAGGGAACCAAGGGTCATTGGGCAAAAACAATTGTTTCTTAAAGCTTT 60

Qy    3706 ttgattaagagaaagaacagaaattttccccaaaagatagtatgaaaaatgaaaaaattt 3765
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTGATTAAGAGAAAGAACAGAAATTTTCCACAAAAGATAGTATGAAAAATGAAAAATTT 120

Qy    3766 gtttcagggcagtagctcagtgtcccccccccttccgaggttccctttgcccccctcaaaa 3825
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTTTCAGGGCAGTAGCTCAGTGTCCCACGTCACTCCTCAACTTTCCTTTCAACCCCTCAAAA 180

Qy    3826 tgttgtatgggttgttctgtaagcagttttgtgaggtgtttaggcaaatgtgtgctgtgtt 3885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGTTGTATGGGGTTACTCTCTAACAGATTCTTGATGTGTCTTTTTCTCTCTATGTGTT 240

Qy    3886 agccttttgttaagatgtaatttttaaatattcagtttatggtactatatttgtaactctc 3945
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGCCTTTGTTAAGATGTAATTTTTAAATATTCAGTTATGGTACTATATTTGTAACTCTC 298


RESULT   5
T38228
ID   T38228 standard; cDNA; 1386 BP.
AC   T38228;
DT   18-DEC-1996  (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA.
KW   Apoptosis-regulating protein; Nip2; adenovirus E1B 19K protein;
KW   Bcl-2; oncogene; cell death; cancer; tumour; immune disorder;
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214. .1158
FT                   /*tag= a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998.
PT   Nucleic acids encoding apoptosis regulating proteins - useful for
PT   diagnosing and treating immune disorders, malignancies, etc.
PS   Claim 1; Page 39-41; 60pp; English.
CC   6 cDNA clones (T38227-32) respectively, each for novel apoptosis-
CC   regulating proteins Nip1, Nip2, Nip3, Bip1A, Bip1 and Bip13
CC   (W00997-W01002).  Nip clones were isolated from human cDNA
CC   libraries using a yeast two-hybrid strategy with adenovirus E1B
CC   19K protein (W01010) as bait; Bip clones were isolated using
CC   Bcl-2 oncoprotein (W01018) as bait.  E1B 19K and Bcl-2 both protect
```

CC cells against apoptosis. Nip1, Nip2, Nip3 bind to 19K, Bip1A and
CC Bip3 bind to 19K and Bcl-2, and Bip5 binds to Bcl-2 alone. Nucleic
CC acids can be used in diagnostic assays, to produce recombinant
CC apoptosis-regulating proteins, or to regulate cell death by
CC increasing or suppressing expression of an apoptosis-blocking gene
CC such as bcl-2.
SQ Sequence 1386 BP; 428 A; 249 C; 336 G; 373 T;

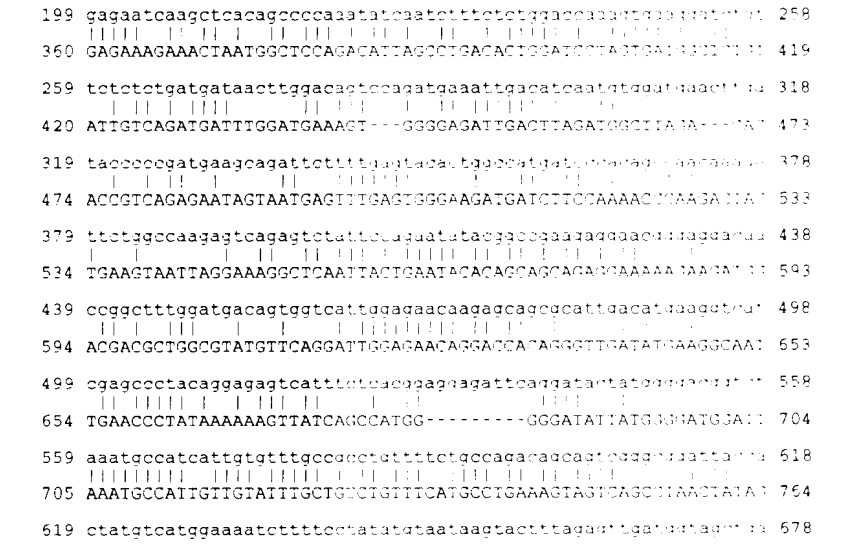
Query Match 6.8%; Score 286.2; DB 1; Length 1386;
Best Local Similarity 61.7%; Pred. No. 2.9e-49;
Matches 513; Conservative 0; Mismatches 303; Indels 15; Gaps 3;

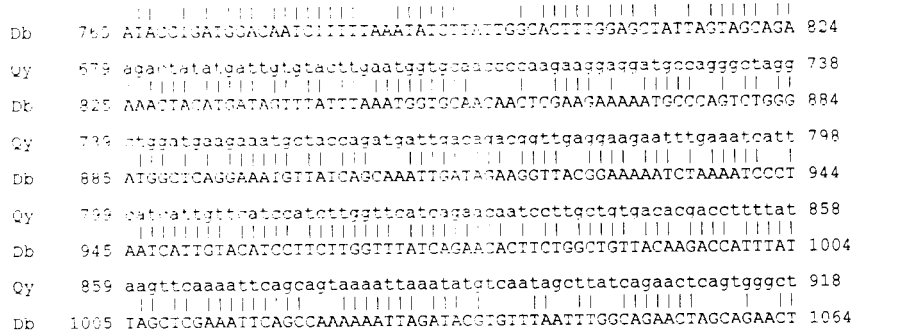


RESULT 5
X04120
ID X04120 standard; cDNA; 1386 BP.

AC X04120;
DT 13-APR-1999 (first entry)
DE Human Nip2 encoding cDNA.
KW Human; Nip1; Nip2; Nip3; Bip1A, Bip5, Bip13, adenovirus, cell death;
KW viral infection; Bcl-2; protooncogene; mutational analysis; apoptosis;
KW E1B 19K protein; cell survival regulation; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 214. .1158
FT /*tag= a
PN US5858678-A.
PD 12-JAN-1999.
PF 21-MAR-1995; 408095.
PR 02-AUG-1994; US-284139
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 99-152099/13
DR P-PSDB; W94324.
PT Polypeptides that bind to anti-apoptotic proteins - useful for
PT protecting against cell death induced by viral infection and
PT modulate response to physical and chemical stimuli
PS Claim 2; Column 51-54; 41pp; English.
CC The present invention describes (1) a method for regulating cell death,
CC comprising exposing an isolated cell to a polypeptide selected from
CC Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13; (2) a method for neutralising
CC the activity of the adenovirus E1B 19 kD protein, the Bcl-2 protein or
CC the BHRF-1 protein, comprising exposing an isolated cell to a
CC polypeptide as in (1); and (3) a method for detecting molecules that
CC bind to at least one polypeptide as in (1), comprising lysing cells,
CC exposing the lysate to the polypeptide and detecting any molecular
CC polypeptide aggregates. The methods are useful for providing proteins
CC able to bind to other proteins known to regulate cell survival e.g. it
CC is known that E1B 19K protein provides a survival function similar to
CC the cellular protooncogene bcl-2 gene product which is able to block
CC apoptosis in haematopoietic B and T cells. The present sequence encodes
CC human Nip2, given in the present invention.
SQ Sequence 1386 BP; 428 A; 249 C; 336 G; 373 T;

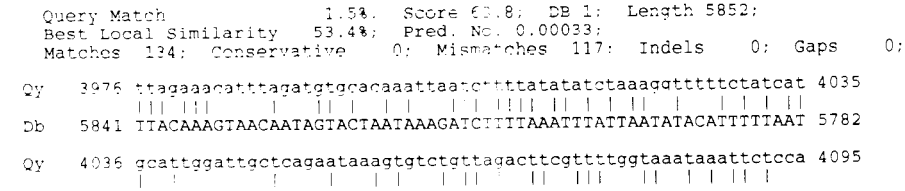
Query Match 6.8%; Score 286.2; DB 1; Length 1386;
Best Local Similarity 61.7%; Pred. No. 2.9e-49;
Matches 513; Conservative 0; Mismatches 303; Indels 15; Gaps 3;





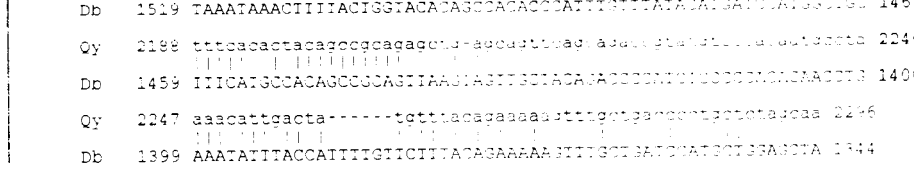
RESULT 7
Q11710/c
ID Q11710 standard; DNA; 5852 BP.
AC Q11710;
DT 30-JUL-1991 (first entry)
DE Dictyostelium plasmid Ddp2 containing Rep gene.
KW slime mould; replication; Rep gene; ss.
OS Dictyostelium discoideum.
FH Key Location/Qualifiers
FT cds 2378. .5041
FT /*tag= a
FT /product= involved in extrachromosomal replication
PN WO9106644-A.
PD 16-MAY-1991.
PF 02-NOV-1990; AU0530
PR 02-NOV-1989; AU-007187
PA (UYMA-) MACQUARIE UNIV.
PI Slade MB, Chang ACM, Williams KL;
DR WPI; 91-164194/22.
DR P-PSDB; R11988
PT Polypeptide facilitating extra-chromosomal replication of
PT recombinant plasmid in Dictyostelium species
PS Claim 15; Fig 1; 90pp; English.
CC The sequence of Ddp2 has been found to contain the putative open
CC reading frame indicated in the Features Table. The possible ORF is
CC flanked by regions with similarity to promoter and poly adenylation
CC signals of known Dictyostelium genes. The RNA and polypeptide
CC product of the Rep gene have not, however, been detected. It is
CC believed that the product is produced in low amounts to positively
CC regulate initiation of plasmid replication. The polypeptide may also
CC contain regions that act as negative regulators of plasmid copy
CC number. See also Q11711 and Q11712.
SQ Sequence 5852 BP; 2298 A; 651 C; 708 G; 2195 T;

Query Match 1.5%; Score 63.8; DB 1; Length 5852;
Best Local Similarity 53.4%; Pred. No. 0.00033;
Matches 134; Conservative 0; Mismatches 117; Indels 0; Gaps 0;



RESULT 8
V44864/c
ID V44864 standard; cDNA; 1536 BP.
AC V44864;
DT 21-OCT-1998 (first entry)
DE Clone CQ331_2 coding sequence.
KW Secreted protein; nutritional source; cell proliferation activity;
KW cell differentiation activity; immune stimulating; tissue g. with activator;
KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
KW tumour inhibitor; clone CQ331_2; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 117. .923
FT /*tag= a
FT /transl_except= (pos:715. .717 aa:Xaa)
FT /note= "Xaa= unspecified amino acid"
PN WO9825952-A2.
PD 18-JUN-1998.
PF 12-DEC-1997; U23224.
PR 11-DEC-1997; US-989232.
PR 13-DEC-1996; US-766263.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, Mccoy JM, Merriens L,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-352424/31
DR P-PSDB; W69245.
PT New isolated polynucleotides - obtained from human adult tissues,
PT human adult ovary, human adult brain and human adult heart cDNA
PT libraries
PS Claim 29; Page 75-76; 108pp; English.
CC This sequence represents a polynucleotide of the invention, and encodes a
CC secreted protein. It was isolated from a human adult heart cDNA library,
CC and is designated clone CQ331_2. The DNA sequences and encoded
CC polypeptides can be used as nutritional sources or supplements, or may
CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
CC immune stimulating or suppressing activity, haematopoiesis regulating
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC activin/inhibin activity, chemostatic/chemokinetic activity,
CC cadherin/tumour invasion suppressor activity, plasma growth activity,
CC tumour inhibition activity or other activities.
SQ Sequence 1536 BP; 471 A; 312 C; 357 G; 415 T;

Query Match 1.5%; Score 62; DB 1; Length 1536;
Best Local Similarity 67.6%; Pred. No. 0.00067;
Matches 119; Conservative 0; Mismatches 51; Indels 7; Gaps 2;



```
RESULT   9
V52604
ID   V52604 standard; cDNA; 2280 BP.
AC   V52604;
DT   21-DEC-1998  (first entry)
DE   Human metalloproteinase BS10.55 cDNA.
KW   Disintegrin metalloproteinase; proteinase, BS10.55, human, ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             61. .1473
FT                   /*tag= a
FT                   /note= "possible initiation at about 61; probable
FT                   initiation at 100"
FT   sig_peptide     61. .150
FT                   /*tag= b
FT   mat_peptide     151. .1470
FT                   /*tag= c
PN   WO9839421-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; 003937.
PR   07-MAR-1997; US-813150.
PA   (SCHE ) SCHERING CORP.
PI   Dowling LM, Gorman DM, Huffine CF, LeBecque SJE,
PI   Liu Y, Mueller CG;
DR   WPI; 98-506353/43.
DR   P-PSDB; W75769.
PT   New isolated genes from human dendritic cells - which encode
PT   products having disintegrin-metalloproteinase monoamine oxidase or
PT   GTPase activities
PS   Disclosure; Page 85-88; 100pp; English.
CC   This cDNA clone codes for human BS10.55.  The BS10.55 gene was
CC   initially found by analysis of clones isolated from germinal centre
CC   dendritic cells.  The translation product (see W75769) exhibits
CC   structural motifs of a member of the disintegrin-metalloproteinase
CC   family.  As such, BS10.55 may be involved in cellular physiology or
CC   development, in dendritic cell antigen processing or presentation,
CC   or in production of important cytokine-like molecules, e.g. Fas
CC   ligand or tumour necrosis factor.  In particular, the protease may
CC   be necessary for the conversion of pro-proteins to proteins, e.g
CC   cytokine or protein precursors to mature forms, or for proper
CC   immunological function, e.g. antigen processing and presentation
CC   The invention provides a process for recombinant production of
CC   BS10.55, host cells, expression vectors and specific antibodies.
CC   The products can be used for detection, diagnosis, development of
CC   therapeutic methods and drug screening.
SQ   Sequence  2280 BP;  758 A;  463 C;  439 G;  620 T;
```

```
  Query Match             1.5%;  Score 61.6;  DB 1;  Length 2280;
  Best Local Similarity   51.2%;  Pred. No. 0.00075;
  Matches  169;  Conservative    0;  Mismatches  159;  Indels    2;  Gaps    1.

Qy   3896 aagatgtaattttaaatattcagttatggtactatgtttgcaactctcgtcttatcacaa 3955
           ||:|  :  | |  | ||  |||||  :  :  ||| |  :  | |  :  ||  | | |
Db   1922 AAGAATAGAATTCAGGTTTTATGTTAATTGCTGTGTTCACTTCGGCCTTGCTCTTGAAAG 1981

Qy   3956 tgcctcagtagtttgttcccttagaaacatttagatgtgcacaaattaatctttttatata 4015
          :|:  :  :  | |  | |  |  |  |        || | | |||  || |||    |
Db   1982 TGCAGTATTTTTCTACATCATGTCGAGAATGATTCAATGTAAATATTTTTCATTTTATCA 2041

Qy   4016 tctaaaggtttttctatcatgcattggattgctcagaataaagtgtctgttagacttcgt 4075
          : || :  | | |  ||  |  |  |        || |||| | | ||       ||
Db   2042 TGTATATCCTATACACACATCTCCTTCATCATCATATATGAAGTTTATTTT--GAGAAGT 2099

Qy   4076 tttggataaataaattctccataatgtagattaataatataaaagtctttaatgacacaat 4135
          :  |  | | | | ||  |  |  |  ||  ||  |||| |||  ||
Db   2100 CTACATTGCTTACATTTTAATTGAGCCAGCAAAGAAGGCTTAATGATTTATTGAACCATA 2159

Qy   4136 atatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaaa 4195
          || || ||| |  | ||  | | |  || ||| |||||  |  |||||||||||
Db   2160 ATGTCAATAAAAACACAACTTTTGAGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2219
```

```
Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||| ||.|||
Db   2220 AAAAAAAAAAAAAAAAAAAAAAGAAAAAA 2249


RESULT  10
V61487
ID   V61487 standard; cDNA; 4237 BP.
AC   V61487;
DT   11-JAN-1999  (first entry)
DE   Human secreted protein fe366_1 cDNA.
KW   Secreted protein; human; fe366_1; ds.
OS   Homo sapiens
FH   Key             Location/Qualifiers
FT   CDS             3746. .4030
FT                   /*tag= a
PN   WO9841539-A2.
PD   24-SEP-1998.
PF   19-MAR-1998; U05474.
PR   18-MAR-1998; US-040963.
PR   19-MAR-1997; US-820493
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D,
PI   Pacie LA, Spaulding V, Treacy M;
DR   WPI; 98-521163/44.
DR   P-PSDB; W79096.
PT   New polynucleotide(s) encoding secreted human proteins  derived
PT   from human foetal kidney, adult testes and adult or foetal brain
PT   cDNA libraries
PS   Claim 36; Page 88-91; 112pp; English.
CC   This full-length cDNA clone, designated fe366_1, codes for a novel
CC   secreted human protein (see W79096).  It was isolated from a human
CC   adult brain cDNA library using methods which are selective for
CC   cDNAs encoding secreted proteins, or was identified as encoding
CC   a secreted or transmembrane protein on the basis of computer
CC   analysis of the encoding protein.  The nucleotide sequence shows
CC   homology to some database sequences, and may contain a CAA repeat
CC   and/or Alu repetitive element.  The invention provides cDNA clones
CC   (see V61477-87) from human foetal kidney, adult testis, and adult
CC   or foetal brain cDNA libraries that code for secreted proteins
CC   (see W79087-97).  These clones are deposited as ATCC 98364.  The
CC   polynucleotides and proteins are predicted to have useful
CC   biological activities which would make them suitable for treating,
CC   preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given.  Suggested
CC   activities include nutritional, immune stimulating (e.g. as
CC   vaccines; or suppressing, haematopoiesis regulating, tissue growth,
CC   activin/inhibin, chemotactic/chemokinetic, haemostatic and
CC   thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour
CC   invasion suppressor and tumour inhibition activities.  The
CC   polynucleotides are also stated to be useful for gene therapy.
SQ   Sequence  4237 BP;  1320 A;  778 C;  781 G;  1358 T;
```

```
  Query Match             1.4%;  Score 60;  DB 1;  Length 4237;
  Best Local Similarity   57.7%;  Pred. No. 0.0018;
  Matches  105;  Conservative    1;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   4044 ttgctcagaataaagtgtctgttagacttcgttttggtaaataaattctccataatgtagt 4103
          || |||  || |||   |   :| |  :  | | || |||| ||     | | |
Db   2384 TTTCTCCCAAAAAATCTTTAAGTAAATATATTTGGGAAAATTAAAAAAAGATTTCAAATA 2443

Qy   4104 attaatatatataaagtctttaatgacacaatatatctatatagcctcactgtataattca 4163
          |  || || | |||      ||  | | || | |  | | |    :  : |.
Db   2444 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2503

Qy   4164 agaaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4223
          | ||| |||||  |      ||||||||||||||||||||||||||||||||||||||
Db   2504 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2563

Qy   4224 aa 4225
          ||
Db   2564 AA 2565
```

```
Db   2564 AA 2565


RESULT  11
V54587
ID   V54587 standard; cDNA; 2447 BP.
AC   V54587;
DT   30-OCT-1998  (first entry)
DE   Human secretory protein encoding cDNA clone C01020-1.
KW   Secretory protein; AM-973-1; BK260-2; BR390-1; CJ539-3; CN729-3;
KW   CJ1390-3; CS1020-1; CS752-3; DM340-1; DW902-1; nutritional source;
KW   cytokine; cell proliferation; immune stimulation; immune suppression;
KW   haematopoiesis regulation; receptor/ligand activity; tumour inhibition;
KW   anti-inflammatory activity; cadherin; tumour invasion suppression;
KW   activin; inhibin; chemostatic; chemokinetic; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             184. .1191
FT                   /*tag= a
FT   CDS             347. .592
FT                   /*tag= b
FT                   /note= "encoded protein shown in W75861"
PN   WO9833916-A2.
PD   06-AUG-1998.
PF   30-JAN-1998; 9801811.
PP   29-JAN-1999; US-014969.
PR   31-JAN-1997; US-792511.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-437469/37.
DR   P-PSDB; W75857, W75861.
PT   New nucleic acid encoding secreted protein from human cells - and
PT   related vectors, transformed cells and proteins
PS   Claim 30; Pages 80-92; 114pp; English.
CC   Sequences shown in V54581 to V54590 represent cDNA sequences encoding
CC   human secretory proteins (W75851 to W75860). The sequences correspond to
CC   the clones designated AM-973-1, BK260-2, BR390-1, CJ539-3, CN729-3,
CC   CJ1390-3, CS1020-1, CS752-3, DM340-1 and DW902-1 respectively. The
CC   polynucleotides and polypeptides can be used as e.g. nutritional sources
CC   or supplements. They may exhibit e.g. cytokine and cell proliferation/
CC   differentiation activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, receptor/ligand activity, anti-
CC   inflammatory activity, cadherin/tumour invasion suppresser activity,
CC   tumour inhibition activity. They may also exhibit other activities such
CC   as tissue growth activity, activin/inhibin activity and chemostatic/
CC   chemokinetic activity.
SQ   Sequence  2447 BP;  932 A;  458 C;  506 G;  551 T;


  Query Match             1.4%;  Score 59.8;  DB 1;  Length 2447;
  Best Local Similarity   52.7%;  Pred. No. 0.0018;
  Matches  127;  Conservative    1;  Mismatches  113;  Indels    0;  Gaps    0;

Qy   3985 tttagatgtgcacaaattaatcttttatatatctaaaggttttttctatcatgcattggat 4044
           | ||||| || | | ||   ||   ||| |  |    || |||    |      ||  |
Db   2161 TATAGATATGTATATATGGTTCACATATTTGCATCTGTGTATTTGATTTTGTACTTTAAA 2220

Qy   4045 tgctcagaataaagtgtctgttagacttcgttttggtaaataaattctccataatgtaga 4104
           ||    ||||||  | ||  |||           |:| |||      | ||   | |
Db   2221 TGTGACAAATAAATTTTTTGGGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2280

Qy   4105 ttaataatataaaagtctttaatgacacaatatatctatatagcctcactgtataattca 4164
            | | | ||||      ||  | | |   |  | | |     |   | ||   |
Db   2281 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2340

Qy   4165 gaaataaaaattgatttctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4224
           ||| |||||   |      ||||||||||||||||||||||||||||||||||||||||
Db   2341 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2400

Qy   4225 a 4225
          |
Db   2401 A 2401


RESULT  12
I90174
ID   I90174 standard; cDNA; 1474 BP.
AC   I90174;
DT   20-APR-1998  (first entry)
DE   Oil seed rape cysteine protease cDNA clone CDCYS61.
KW   Promoter; oil seed rape; cDNA clone CDCYS61;
KW   cysteine protease; disrupter protein; plant; tolerance; herbicide;
KW   insect pest; fungal disease; improved yield; improved quality;
KW   fertility control; ss.
OS   Brassica napus.
FH   Key             Location/Qualifiers
FT   CDS             33. .1157
FT                   /*tag= a
PN   WO9735983-A2.
PD   02-OCT-1997.
PF   18-MAR-1997; G00729.
PR   22-MAR-1996; GB-006062.
PA   (ZENE ) ZENECA LTD.
PI   Greenland AJ, Jepson I, Thomas DRP;
DR   WPI; 97-489646/45.
DR   P-PSDB; W27446.
PT   Cysteine protease promoter - obtained from oil seed rape useful to
PT   develop plants with improved agronomic characteristics
PS   Claim 23; Fig 17; 137pp; English.
CC   The present sequence encodes an oil seed rape cysteine protease,
CC   the promoter for which can [illegible] the expression of a disrupter
CC   protein gene to a suitable stage of plant development, to provide
CC   plants with novel agronomic features, e.g. tolerance to herbicides,
CC   insect pests and fungal diseases, improved yield and/or quality of
CC   harvested product, and novel mechanisms for the control of plant
CC   fertility.
SQ   Sequence  1474 BP;  528 A;  265 C;  316 G;  365 T;


  Query Match             1.4%;  Score 59.6;  DB 1;  Length 1474;
  Best Local Similarity   61.7%;  Pred. No. 0.0017;
  Matches   95;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy   4072 tcgttttggtaaataaaattctc[illegible] 4131
           || |||| | | ||  ||   |  |
Db   1301 TCATTTCGTTTCAAAGAAAGATTCATTTCGTTCTTTTTCTTCTTTAATATTTTATTAAAA 1360

Qy   4132 caatatatctatatagcctcactgtataattcagaaataaaaattgatttctgcaaaaaaa 4191
           |||||| | | |   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   1361 TGATATATTTTTTTTAAAAAAATTTTTTTTTAATAATTTTTTTTTTTTTTTTTTTTTTT 1420

Qy   4192 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4221
          |||||||||||||||||||||||||||||
Db   1421 AAAAAAAAAAAAAAAAAAAAAAATAATAAAAAAAA 1454


RESULT  13
Q04690
ID   Q04690 standard; cDNA; 1834 BP.
AC   Q04690;
DT   05-OCT-1990  (first entry)
DE   Encodes Mammalian amino acid dehydrogenase activating factor-beta
KW   Mammalian amino acid dehydrogenase activating protein-beta;
KW   tyrosine dehydrogenase; tryptophan dehydrogenase; ss.
OS   synthetic.
FH   Key             Location/Qualifiers
FT   polya_signal    1671. .1676
FT                   /*tag= a
FT   cds             166. .1006
FT                   /*tag= b
PN   JP02111796-A.
PD   24-APR-1990.
PF   21-OCT-1988; 264097.
PR   21-OCT-1988; JP-264097.
PA   (JOEN) Joa Nenryo Kogyo KK.
```

```
DR   WPI; 90-169110/22.
DR   P-PSDB; P05084
PT   Amino acid hydrogenase activating protein -
PT   isolated from mammal brain and activates tyrosine hydrogenase
PT   and tryptophan hydrogenase
PS   Disclosure; p; Japanese
CC   Probable error in specification at posn 168, C residue should read
CC   G (to give Met codon ATG instead of ATC).
CC   Encodes a protein characterised by Mr of 27KD (by SDS-PAGE), iso-
CC   electric point of 4.6 and an N-terminal amino acid sequence beginning
CC   MetGlyAsp-.
CC   Protein activates Tyr hydrogenase and Trp hydrogenase in the presence
CC   of calcium ions and calmodulin-dependent protein phosphatase Type II.
CC   Useful as a pharmaceutical analysis reagent.
CC   See also Q04691 and Q04692
SQ   Sequence  1834 BP;  546 A;  454 C;  471 G;  363 T;


  Query Match             1.4%;  Score 59;  DB 1;  Length 1834;
  Best Local Similarity   56.4%;  Pred. No. 0.0024;
  Matches  110;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy   4031 atcatgcattggattgctcagaataaagtgtctgttagacttcgttttggtaaataaattt 4090
           || ||| | ||     ||| ||   |     | |   | |  ||      | | ||||
Db   1621 ATTATGAAATGCCAGACTCCTAAATTAAATGTTTGGAATTCAGTGGGTAAATAAAAATG 1680

Qy   4091 ctccataatgtagattaataatataaaagtctttaatgacacaatatatctatatagcct 4150
           || | |   | | ||||| || | ||||     ||  | | || | |   | | |
Db   1681 CTGCTTTGGGGATATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1740

Qy   4151 cactgtataattcagaaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaa 4210
            |    | ||   | ||| |||||   |      ||||||||||||||||||||||||
Db   1741 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1800

Qy   4211 aaaaaaaaaaaaaa 4225
          ||||||||||||||
Db   1801 AAAAAAAAAAAAAA 1815


RESULT  14
Q58669
ID   Q58669 standard; cDNA; 1172 BP.
AC   Q58669;
DT   17-OCT-1994  (first entry)
DE   Protein allergen of Der pI
KW   Der pI  House Dust Mite Allergen; as
OS   Dermatophagoides pteronyssinus.
FH   Key             Location/Qualifiers
FT   signal_peptide  1. .344
FT                   /*tag=  a
FT                   /product= partial signal peptide
FT                   /note= "344 base long 5'proximal end sequence "
FT   mat_peptide     345  1009
FT                   /*tag=  b
FT                   /product= Der pI mature protein
FT                   /note= "encoded in gull pI"
FT   polya_signal    1078. .1083
FT                   /*tag=  c
PN   WO9405790-A.
PD   17-MAR-1994.
PF   10-SEP-1993; WO08518
PR   10-SEP-1992; US-943299
PA   (IMMU-) IMMULOGIC PHARM CORP.
PI   Chua K;  Thomas WR;
DR   WPI; 94-101195/12.
DR   P-PSDB; P49920
PT   New protein allergens of house dust mite - used for diagnosing
PT   and treating sensitivity in an individual to house dust mite
PT   allergens
FS   Example 1; Fig 21; 98pp; English.
CC   Q58669 is a cDNA clone encoding a full length preproenzyme Der pI.
CC   This is encoded in a single open reading frame, with a termination
CC   site at position 1011-1013. The mature protein can be used to
CC   detect sensitivity in an individual to house dust mite and to reduce
CC   the sensitivity of the individual.
SQ   Sequence  1172 BP;  417 A;  227 C;  197 G;  331 T;


  Query Match             1.4%;  Score 58.8;  DB 1;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 0.0024;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

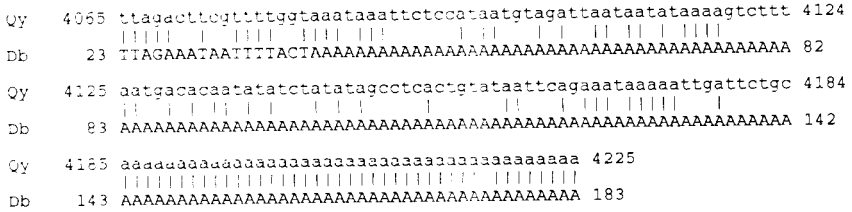
Qy   4076 tttggtaaataaaattctccataatgtagattaataatataaaagtctttaatgacacaat 4135
           ||   |||| |||    | ||    || || | |   | ||  | ||||   | |||
Db   1005 ITCTCTAAACAAAAAGACAATTTCTTAATATCATTGTCACTAATTTATTTAAAATAAAAA 1064

Qy   4136 atatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaaaa 4195
           | | | |   | | | |   ||     | || ||||||   |      ||| ||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCTAAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT  15
T76782
ID   T76782 standard; DNA; 240 BP.
AC   T76782;
DT   15-SEP-1997  (first entry)
DE   Staphylococcus aureus exfoliative toxin A gene capture probe.
KW   Asymmetric polymerase chain reaction; nucleic acid amplification;
KW   PCR; detection; assay; exfoliative toxin A; ETA; skin lesion;
KW   competitive primer; capture probe; ss.
OS   Synthetic.
PN   US5627054-A.
PD   06-MAY-1997.
PF   05-APR-1996; 628417.
PR   05-APR-1996; US-628417.
PA   (USSA ) US SEC OF ARMY
PI   Gillespie D;
DR   WPI; 97-271311/24.
PT   Quantitative nucleic acid amplification - by competitor primer
PT   asymmetric polymerase chain reaction
PS   Example 1; Column 6; 9pp; English.
CC   In a specific example of a novel process for amplifying an amount
CC   (known or unknown) of a double-stranded nucleic acid segment to produce
CC   single-stranded nucleic acid in an amount that is proportional to the
CC   starting amount of the nucleic acid, the Staphylococcus aureus
CC   exfoliative toxin A (ETA) gene was used as the DNA template. The
CC   region comprising nucleotides 105-436 was amplified by symmetric,
CC   asymmetric or competitor primer asymmetric PCR using the primers
CC   ETA-A2 and ETA-B (see T76778 and T76779). For asymmetric PCR, the
CC   amount of primer ETA-B was reduced and for competitor primer
CC   asymmetric PCR a competitor primer ETA-B2 (see T76780) was added
CC   with upstream primer ETA-A2 after the initial cycling reaction. PCR
CC   products containing ETA-specific sequences were detected
CC   radioactively by a capture system which employed a bifunctional
CC   capture probe ETA-CP (see T76781 and T76782). ETA-CP was designed
CC   to capture the amplified sense strand onto capture membranes
CC   through hybridisation between the first 40 nucleotides of ETA-CP
CC   and nucleotides 321-360 of the ETA gene and through hybridisation
CC   of the poly(dA) tail on ETA-CP with poly(dT) tails on the capture
CC   membranes. A radioactively labelled "label probe" (see T76783),
CC   complementary to nucleotides 333-413 of the ETA gene was used to
CC   detect the amplicons. Results showed that hybridisation of the
CC   capture probe and label probe to the denatured symmetric PCR
CC   product was much less efficient than hybridisation to the
CC   single-stranded PCR products of the asymmetric and competitor
CC   primer asymmetric reactions.
SQ   Sequence  240 BP;  221 A;  3 C;  4 G;  12 T;


  Query Match             1.4%;  Score 58.6;  DB 1;  Length 240;
  Best Local Similarity   60.2%;  Pred. No. 0.0019;
  Matches   97;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;
```



Search completed. February 8, 2000, 20:44:06
Job time: 894 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2000, 19:32:30 ; Search time 3770.17 Seconds
(without alignments)
1114.622 Million cell updates/sec

Title: US-09-174-937-3
Perfect score: 1113
Sequence: 1 atggggaccaccgaagccac..... ..atcaggaaacaagcatgtcc 1113

Scoring table: IDENTITY_NUC

Searched: 4038634 seqs, 1887931982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	349.8	31.4	368	20	[illegible]	[illegible] EST03534 Fe
2	269	24.2	427	20	[illegible]	[illegible] AA126572 In
3	240	21.6	499	46	AI437189	AI437189 ib38g06.y
4	181.8	16.3	186	20	[illegible]	[illegible] EST44614 Hu
c 5	141	12.7	461	46	AI120638	AI120638 AA240932
6	136.4	12.3	534	60	AI788178	AI788178 wi18c08.y
7	136	12.2	368	33	AA455831	AA455831 wr55e03.r

```
c    8      126    12.2     495   48  AI620433                       AI620433 tu47b05.x
     9      135    12.1     423   20  T31973                         T31973 EST41767 Hu
    10   132.4    12.0     449   29  AA137877                       AA137877 mq75d03.r
    11   130.8    11.8     516   40  AA914577                       AA914577 wz07e07.r
    12   130.2    11.7     312   20  Z32816                         Z32816 HHEA02P Atr
    13   127.4    11.4     452   28  AA067189                       AA067189 mm38b12.r
    14   127.4    11.4     443   28  AA068579                       AA068579 mm45e06.r
    15   125.6    11.3     810   43  AU035341                       AU035341 AU035341
    16   124.6    11.2     559   79  FR0014610                      AL005822 F.rubripe
    17   109.2     9.8     859   66  AU079281                       AU079281 AU079281
    18   104.4     9.4     392   28  AA113114                       AA113114 zn69b02.r
    19   100.8     9.1     441   21  T94657                         T94657 yo27e08.r1
    20    99.4     8.9     281   20  T31993                         T31993 EST41923 Hu
c   21      99     8.9     526   43  AI180114                       AI180114 EST223851
    22    92.4     8.3     294   29  AA129380                       AA129380 zn84e05.r
    23    88.8     8.0     478   40  AA980800                       AA980800 ua45g06.r
    24    86.6     7.8     357   41  AI006991                       AI006991 ua78a04.r
    25    85.2     7.7     275   34  AA504065                       AA504065 nh40g09.s
c   26    79.2     7.1     378   61  AI852423                       AI852423 UI-M-BH0-
    27    77.4     7.0     457   31  AA280204                       AA280204 zt05e05.r
    28    76.9     6.9     677   69  AW106241                       AW106241 um25a08.y
    29    73.4     6.6     507   28  AA096641                       AA096641 mo09f08.r
    30    73.4     6.6     467   33  AA396890                       AA396890 mr38g06.r
    31    73.4     6.6     557   49  AI663516                       AI663516 uk33g11.y
    32    73.4     6.6     474   62  AI876404                       AI876404 uk74f05.y
    33    73.4     6.6     526   62  AI882015                       AI882015 ul30c05.y
    34    73.4     6.6     524   62  AI931278                       AI931278 ul67b04.y
    35    73.4     6.6     593   64  AW044733                       AW044733 um13h05.y
c   36    69.4     6.2     551   29  AA129317                       AA129317 zn84e05.s
c   37    68.2     6.1     289   35  AA578186                       AA578186 nl29e10.s
    38      63     5.7     212   72  AV293206                       AV293206 AV293206
    39    62.4     5.6     918   79  CNS00AJ3                       AL055957 Drosophil
    40      61     5.5     452   21  T68254                         T68254 yc40g01.r1
    41    60.6     5.4    1101   78  CNS00CB1                       AL071659 Drosophil
    42      59     5.3     569   39  AA866746                       AA866746 ud07e04.r
    43    57.8     5.1     604   51  AI722140                       AI722140 fd19g03.y
    44    57.2     5.1     165   64  AW060351                       AW060351 UI-M-BH1-
    45    56.6     5.1     382   24  D61567                         D61567 HUM418G12B
```

ALIGNMENTS

```
RESULT   1
T05645
LOCUS       T05645         368 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST03534 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDD71, mRNA sequence.
ACCESSION   T05645
VERSION     T05645.1  GI:316795
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):82344"
                     /db_xref="taxon:9606"
                     /clone="HFBDD71"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: Lambda2AP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambda2AP-II
                     vector, 1.0kb average inser size."
BASE COUNT       88 a    106 c     90 g     80 t      4 others
ORIGIN
```

```
  Query Match           31.4%;  Score 349.8;  DB 20;  Length 368;
  Best Local Similarity   98.1%;  Pred. No. 4e-72;
  Matches  361;  Conservative    1;  Mismatches    6;  Indels    1;  Gaps    1;

Qy     615 cgactaccactacatcatggagaacctcttcctgtaccatcatcaacagcttcaacatcc 674
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGACTACCACTACATCATGGAGAACCTCTTCCTGTACCATCATCAACAGCTTCAACATCC 60

Qy     675 ggtggctgaggactacatgatcatcaccgagctgaacaccgtggacaacgacgtcgagtac 734
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGTGGCTGAGGACTACATGATCATCACCGAGCTGAACACCGTGGACAACGACGTCGAGTAC 120

Qy     735 tggaatcggctggctgaagaagtgctaccagatgatcgaccgcgagatgtgcgtcaaggac 794
           |||||||||| ||||||||||||||||||||||||||||||||||| |||||||||| |||
Db     121 TGGAATCGGCTGGCTGAAGAAGTGCTACCAGATGATCGACCGCGA-GTTCGCCAAAAAGCTT 179

Qy     795 gaagtccttgatcatcgtccaccccttgtttcattcgggactgtggtggtgcgtatcttcat 854
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 GAAGTCCTTGATCATCGTCCACCCCTCGTGGTTCATTCGGACTGTGGTGGTGCGTATCTTCAT 239

Qy     855 ccctttcatcagcgtcaagttcatccaagagatccagtacgtgattgccgtgttcgtcgcc 914
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 CCCTTTCATCAGCGTCAAGTTCATCCAAGAGATCCAGTACGTGATCGCCGTGTTCGTCGCC 299

Qy     915 ggagcaactcatcctatggaagcagtcgagatccgagctgcgtcctgcgaagtacgaaga 974
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GGAGCAACTCATCCCTATGGAAGCAGTCGAGATCCGAGCTGCGTCCTGCGAAGTACGAAGA 359

Qy     975 ggaaagact 983
           |||||||||
Db     360 GGAAAGACT 368
```

```
RESULT   2
T08680
LOCUS       T08680         437 bp    mRNA            EST       03-AUG-1993
DEFINITION  EST05572 Infant Brain, Bento Soares Homo Sapiens cDNA clone HIBBI25
            5' end, mRNA sequence.
ACCESSION   T08680
VERSION     T08680.1  GI:389708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..437
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):65191"
                     /db_xref="taxon:9606"
                     /clone="HIBBI25"
```

```
                     /clone_lib="Infant Brain, Bento Soares"
BASE COUNT      117 a    127 c    107 g     84 t      2 others
ORIGIN


  Query Match             24.2%;  Score 269;  DB 20;  Length 437;
  Best Local Similarity   100.0%;  Prod. No. 3.5e-53;
  Matches  269;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     845 ccatctctcgcccttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagct 904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCATCTCTCGCCCTTTCATCAGCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCT 60

Qy     905 tggaagacctggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgc 964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGGAAGACCTGGAGCAACTCATCCCTATGGAACACGTCCAGATCCCAGACTGCGTCCTGC 120

Qy     965 aatacgaagaggaaagactgaaggccaggagggagagcgcgaggccccagccggagtttg 1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATACGAAGAGGAAAGACTGAAGGCCAGGAGGGAGAGCGCGAGGCCCCAGCCGGAGTTTG 180

Qy    1025 tgctgcccaggtctgaagagaagccagaggtggcaccagtggaaaacaggtctgctctgg 1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTGCCCAGGTCTGAAGAGAAGCCAGAGGTGGCACCAGTGGAAAACAGGTCTGCTCTGG 240

Qy    1085 tctcagaagatcaggaaacaagcatgtcc 1113
           |||||||||||||||||||||||||||||
Db     241 TCTCAGAAGATCAGGAAACAAGCATGTCC 269


RESULT   3
AI437189
LOCUS       AI437189       499 bp    mRNA            EST       09-MAR-1999
DEFINITION  fb38g06.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
            TR:Q12982 Q12982 NIP2. , mRNA sequence.
ACCESSION   AI437189
VERSION     AI437189.1  GI:4288571
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,P., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,P.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138723.
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark  DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank  Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1..499
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; cDNA was Not I-linker/adaptor Sal I; 1st
                     strand cDNA was primed with a Not I-oligo(dT)18 primer
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTTT-3');
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab), Jeff Lee, Lucy Li and Max Planck
                     Institut fuer Molekulare Genetik Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      163 a    110 c    117 g    109 t
ORIGIN


  Query Match             21.6%;  Score 241;  DB 4;  Length 449;
  Best Local Similarity   74.3%;  Prod. No. 2.2e-46;
  Matches  303;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    0;

Qy     612 ccccgactaccactacatcatgcagaacctgttcctgtacatcatcaacaactacagct 671
           ||| || |||| ||| | |||| | |  |  |  |  |  |  |  |  |  |  |||
Db       2 CCCAGATTACCATTACCTATGCAAAAT-TGTTTCTGTATTTTCAAGTAGTTTCCACAT 61

Qy     672 cctactgactgaggactacatgatctgtgtgtcctgcgaggtcatcaacaacgacgtgtac 731
           ||| | ||| .||'|| | |  |  |  |  |  |  |  |  |  |  |  |  |||
Db      62 GCTGGTGGCCGCAGGATTACTTCATCATCTTACTTCAATCCATCAGCAAGCCCCCCTAAAAT 121

Qy     732 gcctggaatcggctggctgaagaagtgctacgagatgatcgacaacgacgttgtggcaaaaa 791
           | | || ||  |||| |  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db     122 GCCCGGCATTAGCTGGCTCAAGAATTTTATCAGATCATCAACAACGACTTAGAAAAAAA 181

Qy     792 cctgaagtccttgatcatcgtcgacccctcgtgttcgattcgaacgtggctgcggccatcc 851
           ||| || ||  |||||| |  |  |  |  |  |  |  |  |  |  |  |  |  |||
Db     182 CCTAAAATCTCTGATCATCACTCATCCCTCCTCGTTTATTCCACAGCCGCATCTCATTTC 241

Qy     852 tgcgcctttcatcaggctgaagttcatcaacctgatccaatacatcgtggacgagttcgaaga 911
           || |||| ||||| || |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     242 CAAACCCTTCATCAGCGTCAAGTTCATCAATAAGATCCTTTACTTCATCAGCTTTCAACA 301

Qy     912 cctggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgcgaatacga 971
           ||||| |  | | || |||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     302 GCTGGAAAAGTTCGTTCCCATGGATCATATTCATATCCCAGAATCGATCTTACAATATCA 361

Qy     972 agaggaaagactgaaggccaggagggagagcgcgaggccccagccggagtttg 1025
           |||||| |  |||| || ||  |  |  |  |  |  |  |  |  |  |  |||
Db     362 AGAGGAACGGTTGAACTCCAGAAAAGAGAAGGAGAGCGAGGCCCCAGCCGGAGTTTG 421


RESULT   4
T32180
LOCUS       T32180         184 bp    mRNA            EST       6-SEP-1995
DEFINITION  EST44614 Human Brain Homo sapiens cDNA 5' end, similar to None, mRNA
            sequence.
ACCESSION   T32180
VERSION     T32180.1  GI:614278
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186)
```

```
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,Jr.,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Rednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., , Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: IHC10609
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..186
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):102551"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Brain"
                     /note="Organ: brain"
BASE COUNT       46 a     57 c     38 g     44 t      1 others
ORIGIN


  Query Match             16.3%;  Score 181.8;  DB 20;  Length 186;
  Best Local Similarity   98.4%;  Pred. No. 6.1e-33;
  Matches  183;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    788 aaaacctgaagtcctttgatcatcgtccaccccttcgtggttcattcggactgtgctggcca 847
          ||||||:|||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db      1 AAAACCTGAAGTCCTTGATCATCGTCCACCCCTCGTGGTTCATTCGGACTGTGCTGGCCA 60

Qy    848 tctctcgccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttgg 907
          |||||||||||||||||||||||||||| ||  ||||||||||||||||||||||||||||
Db     61 TCTCTCGCCCTTTCATCAGCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCTTGG 120

Qy    908 aagacctggagcaactcatccctatggaacacctccagatcccagactgcgtcctgcaat 967
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGACCTGGAGCAACTNATCCCTATGGAACACCTACAGATCCCAGACTGCGTCCTGCAAT 180

Qy    968 acgaag 973
          ||||||
Db    181 ACGTAG 186


RESULT   5
AI412633/c
LOCUS       AI412633      461 bp    mRNA            EST       09-FEB-1999
DEFINITION  EST242382 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRDK91 3' end, mRNA sequence.
ACCESSION   AI412633
VERSION     AI412633.1  GI:4256137
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (RPST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3086594.
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..461
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RBRDK91"
                     /clone_lib="Normalized rat brain, Bento Soares"
                     /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT       97 a    118 c    130 g    116 t
ORIGIN


  Query Match             12.7%;  Score 141;  DB 46;  Length 461;
  Best Local Similarity   75.1%;  Pred. No. 3.1e-23;
  Matches  190;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    1;

Qy    864 cagcgtcaagttcatcaacaagatccagtacgtgcacagcttggaagacctggagcaa 923
          ||| |||||||||||||  || ||| |||| ||||||||||| || || |||||||| ||
Db    461 CAGTGTCAAGTTCATCAGTAAAATCCAGTATGTGCACAGTTTGGAGGAGTTGGAATGGAA 402

Qy    924 catccctatggaacacgtccagatcccagactgcgtcctgcaatacgaag 973
          |||||| ||||| ||||| ||| | |||| ||| ||| |||||| | |||
Db    401 CATCCCCATGGAGCACGTGCAGCTGCCAGCCTGTGTCATGCAAATATGACGAATGGAATGT 342

Qy    974 gaaggccaggagggagagcgcgaggc---ccccagccggaagtttgtgctgcccaaggtctaa 1030
          |||| ||||||||||||| ||  || |||||  |||| | |||||| ||||
Db    341 CCGAGCCATGAGGGAGAGCGCTCGGCCAACCGCAGCCCGAACTCCTTTGCTTAGGTCTGA 282

Qy   1031 agagaagccagaggtggcaccagtgcaaaaacaggtctgctctggtctcaggaatctcaaga 1090
          ||| || ||||||   |   || ||  |||| | |  ||  |||| |  |  |  |  |
Db    281 AGAAAAACCAGAGACTGTCGAACAAGACGACAGGGCCCAAGACGTGACAGATTCCCCTA 222

Qy   1091 aacaagcatgtcc 1113
          ||| |||||||||
Db    221 AACGAGCATGTCC 209


RESULT   6
AI788173
LOCUS       AI788173      534 bp    mRNA            EST       02-JUL-1999
DEFINITION  ul18c08.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:2087912 5' similar to TR:Q12982 Q12982 NIPSNAP      mRNA
            sequence.
ACCESSION   AI788173
VERSION     AI788173.1  GI:5335889
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Shurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., Meluzm,R.,
```

```
            Waterston,R. and Wilson,R
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133758.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway  Box 8501, St  Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:995602
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2087918"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dI) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb   Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTGGAGCAGA "
BASE COUNT      118 a    156 c    165 g     95 t
ORIGIN


  Query Match             12.3%;  Score 136.4;  DB 60;  Length 534;
  Best Local Similarity   81.4%;  Pred. No. 4e-22;
  Matches  158;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 atgggaccaccgaagccacggctccggatggaaaacgtggacgtgaaggaggaatggcag 60
           ||||| ||||| |||||  |  ||  |||||||||| |||||||||| ||| |||||||||
Db     340 AIGGGAACCACAGAAGCIACACIAAGGATGGAAAATGTGGACGTGAGGGATGAATGGCAG 399

Qy      61 gacgaagatcttcccaggccactcccagaagagacgggggtggaactgcttggcagcccg 120
           || || ||||| ||||| || |||||||| || || |||||||| | ||| ||  || |
Db     400 GATGAGGATCTGCCCAGACCGCTCCCACAAGACACCGGGGTGGAGCGGCTGGGTGGCGCA 459

Qy     121 gtggaaagacacatcctctcctcccaacacgctaaatttcaacggagcgcatcgtaagagg 180
           ||||||||| | ||||| |||||  ||| | || || | |||||| ||||| |||||
Db     460 GTGGAAGACTGCTGCTCACCTCCCTCCACCCTGAACTTGAGCGGAGCACATCGAAAGAGA 519

Qy     181 aagacgctggtggc 194
           ||||||||||||||
Db     520 AAGACGCTGGTGGC 533


RESULT   7
AA450831
LOCUS       AA450831      368 bp    mRNA            EST       04-JUN-1997
DEFINITION  vg55e03.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865276 5' similar to TR:G559944 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA450831
VERSION     AA450831.1  GI:2164501
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dunoque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Mitreva,M.,
            Schellenberg,K., Stekiel,M., Tan,F., Thiruvilla,R., Moore,R.,
            Ineising,B., Wylie,T., Beddington,R., Stubbs,L., Watson,M. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1147446.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63110
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509364
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /strain="C57BL6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:865276"
                     /clone_lib="Beddington mouse embryonic region"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT. Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings, excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                     Referenced in Development 121, 2479-2494 (1995)"
BASE COUNT      110 a     66 c     98 g     94 t
ORIGIN


  Query Match             12.2%;  Score 134;  DB 33;  Length 368;
  Best Local Similarity   65.5%;  Pred. No. 4.2e-22;
  Matches  199;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     470 tgatcatcgggagcaagaagcgacgctatagacctgcacatgatgtcgccttacatcaaaag 529
           |:|  |: || ||||| |:| |                                      |
Db      58 TCAGGATTGGAGAGCAGGACCACAAGGGTGGACATGTACCCGATTGACCCCTATAAAAAAG 117

Qy     530 tggtcacccacggaagggtactatggcgaaggcctcaacgcgatcatcgtcttcgaggcct 589
           |:: : ||| ||||| |: |                                      |
Db     118 TTATTAGCCATGGAGGATATTATGGGGACGGCTTAAATGCCATTCCTTATTTGTTGTTGT 177

Qy     590 gcttcctccagaacagcagcctacccgactgcaactacatcatatagaagcctatgctat 649
           | |||| | || || ||                                      |
Db     178 GTTTCATGCCIGAGAGIGGTCAGCCTAACTATAGATATTTGATGAAAATCTTCCTAAAA 237

Qy     650 acgtcatcagcagcttagaggctcctggatgactgagtactacatgatgtctgaatg 709
           |:  ||  ||| | |:|:|                                      |
Db     238 ATGTTATTGGCACTTTAGAGCCTTTCCTTACCACACAGTTCTTTGTTTTATTTAAAATG 297

Qy     710 gtgccacgcccccggcgagatcctttaatcggctggctgaatgaagtgctatgcaaatga 769
           |||  ||  | ||  | |                                      |
Db     298 GTGCGACAACTCGAAGAAAAATGCTTAGTTTGGGATGCCTCTGTAGATCCTTATGTAAA 357

Qy     770 tcga 773
           ||||
Db     358 TTGA 361
```

```
RESULT   8
AI620433/c
LOCUS       AI620433      486 bp    mRNA            EST       21-APR-1999
DEFINITION  [illegible] NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE 2254161 3'
            similar to TR:Q12992 Q12992 NIP2. ;, mRNA sequence.
ACCESSION   AI620433
VERSION     AI620433.1  GI:4629559
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949357.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40^UP from Gibco
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2254161"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-989759, 1101161-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      119 a    117 c    120 g    130 t
ORIGIN
```

```
  Query Match           12.2%;  Score 136;  DB 48;  Length 486;
  Best Local Similarity   57.5%;  Pred. No. 4.8e-22;
  Matches  244;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy     544 gagtactacggcgaaggcctcaagcgcatcatgtcttcgcagcctgcttccttccagac 603
           |||||| ||   | |||||||| ||    ||| | || || ||  |||| |  || ||| |
Db     486 GGGTACCCCGTGAAGGGCTCAAGGCGTCAATCCTTTTGGCTTCCTGTTATCTACCAAAA 427

Qy     604 agcagcctcccgcgactaccactacatcatggagaacctcttcctgtacgtcatcagcagc 663
           ||  |||| | |   |  ||| ||  ||  ||  ||  ||| |  | | | |
Db     426 GCAAGAATCCCCACCACCCTTTGTGTCATGGAACACTGGTTTAGGTATATGGTGGGAACT 367

Qy     664 ttagaactcctgttgctgaggactacatgatgtgtacctgaacggtgccacgcccgg 723
           |||| |  | |||||  | |||  |  ||  |  | |  || | |||   | ||
Db     366 CTGAAGCTGCTAGTAGCTGAAAATTACCTGCTTGTTCATTTGAGTGGAGGCACAAGCAGG 307

Qy     724 cgaaggatgcctggaatcggctgctgaaaaatactaccagatgatcgaccggaggttg 782
```

```
Db     306 GCCCAAGTTCCACCTTAAAGCTGGATACGTCAGTGTTACCGTACCCTTAATCTTTTTA 247

Qy     784 cggaaaaacctgaagtccttgatcatcctcatcctgttgttcatt[illegible]tttct 843
           ||||||||||||   || || |  | ||  |  | |||  |
Db     246 CGGAAAAACCTGCGAGCCCTGGTCCTTCTCATGCTACATCGTATCTGAAACTATTCCT 187

Qy     844 gccatctctgcccttttcatcagcgtccaagttcatcaacaagatccaatacgtgtgtgac 903
           ||  |  ||| || ||||||||  ||| ||||    ||  |||  |  |  |  |
Db     186 GCACTGCTTCGGCCCTTCATCAGTTTCAAATTCACACGAAAAATCCGTTTTCTTGACAGT 127

Qy     904 ttggaagacctggagcaactcatccatatggaacacgtccagatcccaaactatctcctg 963
           |||  || ||||  |||||||'     |||| || ||||| ||||| ||    |
Db     126 CTGGGGGAGCTGGCCCAACTCATATCCTCGATCAAGTCCAGATCCGTCAAGTATTAAA 67

Qy     964 caat 967
           || |
Db      66 CAGT 63
```

```
RESULT   9
T31973
LOCUS       T31973        423 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST41767 Human Uterus Homo sapiens cDNA 5' end similar to similar
            to rhoGAP protein, mRNA sequence.
ACCESSION   T31973
VERSION     T31973.1  GI:614071
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fuhrmann,J.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,Jr.,P.S.,
            Kelley,J.M., Klimek,K.M., Kolluri,J.C., Liu,L.-I., Marrakos,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,J., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expressed Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC14305
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd. Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..423
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):102311"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Uterus"
                     /note="Organ: uterus"
BASE COUNT      128 a     71 c     90 g    131 t      3 others
```

```
ORIGIN


  Query Match             12.1%;  Score 135;  DB 20;  Length 423;
  Best Local Similarity   61.9%;  Pred. No. 7.7e-22;
  Matches  244;  Conservative    0;  Mismatches  148;  Indels    2;  Gaps    2;

Qy     517 ccttacatgaaagtggtcacccacggagggtactacggcgaaggcctcaacgccatcatc 576
           || || |  ||'|  ||| ||| || || |  ||  | || ||  | || |||||  |
Db      11 CCCIA!AAAAAAGIIAICAGCCAIGGGGGAIAIIAINGGGAIGGAIIAAAIGCCAIIGII 70

Qy     577 gtcttcgcagcctgcttccttccagacagcagcctccccgactaccactacatcatggag 636
           || || || | ||| ||| | || || || || |  ||  ||||    ||| | |||||
Db      71 GIGIIIGCIGICIGIIICAIGCCIGAAAGIAGCCAGCCTAACIATAGAIACCTGAIGGAC 130

Qy     637 aacctcttcctgtacgtcatcagcagcttagagctcctggtggctgaggactacatgatc 696
           || || ||    || || ||  |||  || ||||  | || || ||  ||||||||||
Db     131 AATCTTTTTAAATATGTTATTGGCACIIIGGAGCTATTAGTAGCAGAAAACTACATGATA 190

Qy     697 gtgtacctgaacggtgccacgccccggcggaggatgcctggaatcggctggctgaagaag 756
           || ||  ||| ||||| ||  | ||  | |  ||||  |  | || ||||| | |||
Db     191 GTTTAIIIAAATGGIGCAACAACTCGAAGAAAAATGCCCAGICTGGGATGGCTCAGGAAA 250

Qy     757 tgctaccagatgatcgaccggaggtttgc-ggaaaaacctgaagtccttgatcatcgtcca 815
           || || |||   || ||  | ||||| | || |||  || || ||| | ||||| || ||
Db     251 IGIIAICAGCAAAIIGAIAGAAGGIIACGGGAAAAATCIAAAATCCCIAAICAIIGIACA 310

Qy     816 ccctcgtggttcattcggactgtgctggc-catctctcgccctttcatcagcgtcaagt 874
           | || ||| | ||  | ||  | ||||   |     |||||| || |||    | |
Db     311 ICCIICIIGGGIIAICAGAACACIICIGGCIGIIANAAGGCCCTTTTATTAGCICGGAAI 370

Qy     875 tcatcaacaaagatccagtacgtgcacagcttgga 908
           |      || ||   |||| |   |     |
Db     371 TICAGCCAAAAATTTGGIACGGGTTTAIIINGGA 404


RESULT  10
AA137873
LOCUS       AA137873      449 bp    mRNA            EST       09-FEB-1997
DEFINITION  mq75d03.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:584549 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA137873
VERSION     AA137873.1  GI:1700244
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394228.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:359197
            Seq primer. -28m13 rev1 ET from Amersham
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:584549"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SURE (tetracycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGATTTTTTTTTTTTTTTTT 3'"
BASE COUNT      147 a     76 c    124 g    102 t
ORIGIN


  Query Match             12.0%;  Score 133.4;  DB 20;  Length 444;
  Best Local Similarity   64.0%;  Pred. No. 1.9e-21;
  Matches  217;  Conservative    0;  Mismatches  121;  Indels    1;  Gaps    1;

Qy     445 gccaacgggcgcctgtggcggacagtgatcatcggaggagcaagatcaccatatagaactg 504
           '  ||||| |||' |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     108 GGCGATGGACGCCGCTGGCCAATGTTCAGGATTGGAGAGCAGGATCACAGGTIGGACATG 167

Qy     505 cacatgatccggccttacatgaccgtggtcacccagggaggctatgtaccgcgaaggcctc 564
           |  ||||  ||| ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     168 AAGGCGATCGAGCCCTACAAAAAAGTTATTAGCCATGGAGGATATATGTGGGACCGCTTA 227

Qy     565 aacgccatcatcg-tcttcgcagctgctgttcctttcagacagcagcctccccgactacca 623
           '| ||||  | | | | || |  |  |  |  |  |  |  |  |  |  |  |  |
Db     228 AATGCCATTGTTGTTATTTGCTTTTCTTTCAATGGTCAGTGTCTTCAGCCTAAGTATAG 287

Qy     624 ctacatcatggagaacctcttcctgtacgtcatcagcagcttagagctcctggtggctga 683
           |'  | ||||| || ||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     288 ATATTTGATGGACAATCTCTTTAAATATGTTATTGTCAGTTTAGAGCTGTTGTTAGCAGA 347

Qy     684 ggactacatgatcgtgtacctgaacggtgccacgccccggcggaggatgcctggaatcggc 743
           | |||'| |||'| | |'| |
Db     348 GAACTACATGATCATTTACTTAAATGCCTCCGACAACTCCAACGAAAAATGCCTAGTTTCGG 407

Qy     744 ctggctgaagaagtgctaccagatgatcgaccggaggtt 782
           |||  | | |   |||'|||  |  |  |  |
Db     408 ATGGCTCAGGAGATGCTACCAGCAAATTGATAGACGATT 446


RESULT  11
AA914577
LOCUS       AA914577      516 bp    mRNA            EST       11-DEC-1998
DEFINITION  wi22g07.r1 Stanford Human Adult (Normal) Muscle Mus musculus cDNA clone
            IMAGE:1314564 5' similar to TR:G12982 G12982 NIP2. ;, mRNA
            sequence.
ACCESSION   AA914577
VERSION     AA914577.1  GI:3053969
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407944.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```

```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:684952
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1..515
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1314564"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      148 a    110 c    118 g    140 t
ORIGIN


  Query Match           11.8%;  Score 138.8;  DB 40;  Length 516;
  Best Local Similarity   69.3%;  Pred. No. 8e-21;
  Matches  192;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     684 ggactacatgatcgtgtacctgaacggtgccacgccccggcggaggatgcctggaatcgg 743
           | ||||||||||| | ||| | || |||||  |  | ||  | |  |||||  |  | ||
Db       1 GAACTACATGATCATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGG 60

Qy     744 ctggctgaagaagtgctaccagatgatcgaccggaggttgcggaaaaacctgaagtccttt 803
           |||||| ||  | ||||||  ||  |    | ||   |||||||| ||||  |
Db      61 ATGGCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTAAGTCTCT 120

Qy     804 gatcatcgtccacccctcgtggttcattcggactgtgctggccatctctcgccctttcat 863
           |||| || ||||||| ||||  ||  |    | ||||  |  | || || || ||
Db     121 AATCATTGTGCACCCCTCTTGGTTTATCCGAACACTTCTGGCTGTTACACGACCATTTAT 180

Qy     864 cagcgtcaagttcatcaacaagatccagtacttgcacagcttggaagacctggagcaact 923
           |||  |||||  | |||  |||  ||  | ||  |||||  || ||||  | ||
Db     181 CAGTTCAAAGTTCAGTCAGAAGATCAGATACTTCTTTAACTTGGCCGAGCTGGCAGAGCT 240

Qy     924 gatcctatggaaacaatccagatccaatctggtcctgcaatagaagagga 977
           | || ||||  | |||  || ||| |  | ||  |||  ||  | ||
Db     241 TGTTCCCATGGAGTATGTCGGCATACCAGAGTTCATCAAACAAGTTGATCAAGA 294


RESULT  12
Z32816
LOCUS       Z32816        312 bp    mRNA            EST       08-MAY-1994
DEFINITION  HHEA02P Atrium cDNA library Human heart Homo sapiens cDNA clone
            HEA02P, mRNA sequence.
ACCESSION   Z32816
VERSION     Z32816.1  GI:483499
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Dupret,S., Hoebeke,P., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE, molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 319 (2), 269-272 (1995)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:307615.
            Contact: Genzentrum Muenchen
            Laboratorium fuer molekulare Biologie
            Am Klopferspitz 18a,8033 Martinsried,Germany
            Email: obermaier@vms.biochem.mpg.de
            single read.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEA02P"
                     /clone_lib="Atrium cDNA library Human heart"
BASE COUNT       97 a     55 c     59 g    101 t
ORIGIN


  Query Match           11.7%;  Score 136.2;  DB 20;  Length 312;
  Best Local Similarity   63.7%;  Pred. No. 9e-21;
  Matches  198;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     571 atcatcgtcttcgcagcctgctttttcagaagaggagctgcccgattacgattgatta 630
           ||  | || |  || | ||| ||  | || ||  |      |  |  |
Db       1 ATTTTTGTGTTTGCTGTCTGTTTCATGCTTGAAAGTAGTCAGCTTAACTATAGATATTT 60

Qy     631 atggagaacctcttcctgtacgtcatcagcagcttagaggctcctggtgactgaggaactg 690
           |||||  || |  ||    || || ||  |||  || |||||  |  || || |    || |
Db      61 ATGGACAATCTTTTTAAATATGTTATTGGCACTTTGGAGCTACTACTAGCAGAAAACTAT 120

Qy     691 atgatcatgtacctgaacggtgccacgccccggcggaggatgcctggaatcggctggct 750
           |||||  || |   | || |||||  ||  | ||  ||  ||  |  |  |  |  |
Db     121 ATGATAGTTTATTTAAATGGTGCAACAACTCGAAGAAAAATGCCCAGTTTGGATGGTTT 180

Qy     751 aagaagtgctaccagatgatcgaccggaggttgcggaaaaacctgaagtccttgatcatc 810
           | | | || || |||   || ||  | ||||| ||| |||| || || |||  | | |
Db     181 AGGTAATGTTATCAGCAAATTGATAGAAGGTTACGGGAAAATCTAAAATCCCTAATTAAT 240

Qy     811 gtccacccctcgtggttcattcggactgtgctggccatctctcgccctttcatcagcgtc 870
           || || || || ||||| ||  | | |  | ||||  |  |  |  |  | |  |
Db     241 GTACATCCTTCTTGGTTTATCAGAACACTTCTGGCTGTTACAAGACCATTTATTAGCTTT 300

Qy     871 aagttcatcaa 881
           || |||| | |
Db     301 AAATTCAGCCA 311


RESULT  13
AA067188
LOCUS       AA067188       452 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm38b12.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:523775 5' similar to TR:O55844 O55844 NTAL-1, mRNA
            sequence.
ACCESSION   AA067188
VERSION     AA067188.1  GI:1564954
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,C., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1323488.
            Contact: Marra M/Mouse EST Project
```

```
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63109
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:317623
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:523775"
                     /clone_lib-"Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host-"SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      145 a     90 c    108 g    109 t
ORIGIN


  Query Match           11.4%;  Score 127.4;  DB 28;  Length 452;
  Best Local Similarity  65.1%;  Pred. No. 4.7e-20;
  Matches  188;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    698 tgtacctgaacggtgccacgcccccggcggaggatgcctggaatcggctggctgaagaagt 757
          | ||| ' || |||| ||  | ||  | '  |||| |  | || |||| | ||  |
Db      2 TTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGATGGCTCAGGAGAT 61

Qy    758 gctaccagatgatcgaccggaggttgcggaaaaacctgaagtcctttgatcatcgtccacc 817
          ||||||||   || ||  |  ||||  || | |||| |||||  | ||||| || ||||
Db     62 GCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTAAGTCTCTAATCATTGTGCACC 121

Qy    818 cctcgtggttcattcggactgtgctggccatctctcgccctttcatcagcgtcaagttca 877
          |||| ||||| || || ||  | |||||  |  | || || || |||||    |||||||
Db    122 CCTCTTGGTTTATCCGAACACTTCTGGCGTGTTACACGACCATTTATCAGTTCAAAGTTCA 181

Qy    878 tcaacaagatccagtacgtgcacagcttggaagacctggagcaactcatccctatggaac 937
          | |||||   |||||    | ||||  || ||||   | ||  | || |||||
Db    182 GTCAGAAGATCAGATACGTCTTTAACTTGGCCGAGTTGGCAGAGCTTGTTCCCATGGAGT 241

Qy    938 acgtccagatcccagactgcgtcctgcaatacgaagaggaaagactgaa 986
          | |||   || ||||| ||| ||   || ||  |||| ||||   | ||
Db    242 ATGTCGGCATACCAGAGTGCATCAAACAGTATGAAGAAGAAAAGTTTAA 290


RESULT  14
AA068579
LOCUS       AA068579      443 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm45e06.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:524482 5' similar to TR:G558944 G558944 NIP2 ., mRNA
            sequence.
ACCESSION   AA068579
VERSION     AA068579.1  GI:1575939
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,P., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:1182468.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63109
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318330
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:524482"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      143 a     89 c    108 g    103 t
ORIGIN


  Query Match           11.4%;  Score 127.4;  DB 28;  Length 443;
  Best Local Similarity  65.1%;  Pred. No. 4.7e-20;
  Matches  188;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    698 tgtacctgaacggtgccacgcccccggcggaggatgcctggaatcggctggctgaagaagt 757
          | ||| ' || |||| ||  | ||  | '  |||| |  | || |||| | ||  |
Db      2 TTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGATGGCTCAGGAGAT 61

Qy    758 gctaccagatgatcgaccggaggttgcggaaaaacctgaagtcctttgatcatcgtccacc 817
          ||||||||   || ||  |  ||||  || | |||| |||||  | ||||| || ||||
Db     62 GCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTAAGTCTCTAATCATTGTGCACC 121

Qy    818 cctcgtggttcattcggactgtgctggccatctctcgccctttcatcagcgtcaagttca 877
          |||| ||||| || || ||  | |||||  |  | || || || |||||    |||||||
Db    122 CCTCTTGGTTTATCCGAACACTTCTGGCGTGTTACACGACCATTTATCAGTTCAAAGTTCA 181

Qy    878 tcaacaagatccagtacgtgcacagcttggaagacctggagcaactcatccctatggaac 937
          | |||||   |||||    | ||||  || ||||   | ||  | || |||||
Db    182 GTCAGAAGATCAGATACGTCTTTAACTTGGCCGAGTTGGCAGAGCTTGTTCCCATGGAGT 241

Qy    938 acgtccagatcccagactgcgtcctgcaatacgaagaggaaagactgaa 986
          | |||   || ||||| ||| ||   || ||  |||| ||||   | ||
Db    242 ATGTCGGCATACCAGAGTGCATCAAACAGTATGAAGAAGAAAAGTTTAA 290


RESULT  15
AU035341
LOCUS       AU035341      810 bp    mRNA            EST       19-OCT-1998
DEFINITION  AU035341 Sugano mouse kidney mRNA No. 6 similar to DNA-binding
            mRNA sequence.
ACCESSION   AU035341
VERSION     AU035341.1  GI:3718344
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 810)
```

```
  AUTHORS   Sasaki,Z., Suzuki,Y., Watanabe,M., Imai,J., Shibui,A., Yoshida,K.,
            Hata,H., Yamaguchi,R., Tateyama,S. and Sugano,S.
  TITLE     Construction of mouse full length-enriched cDNA libraries
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2045191.
            Contact: Katsuyuki Hashimoto
            Division of Genetic Resources
            National Institute of Infectious Diseases
            23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
            Email: khashi@nih.go.jp.
FEATURES             Location/Qualifiers
     source          1. .810
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="MNCb-0342"
                     /clone_lib="Sugano mouse brain mncb"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="TOP10"
                     /note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT] double-stranded cDNA was
                     ligated to a DraIII adaptor [TGTTGGCCTACTGG] , digested
                     and cloned into dstinct DraIII sites of the pME18S-FL3.
                     XhoI sites just outside the DraIII sites can be used to
                     isolate the cDNA insert.  Size selection was performed to
                     exclude fragments <1.5 kb. Library was constructed by
                     Sugano et al. (University of Tokyo, Institute of Medical
                     Science).  Custom primer for sequencing: 5' end primer
                     [CTTCTGCTCTAAAAGCTGCG]"
BASE COUNT      198 a    206 c    233 g    156 t     17 others
ORIGIN


  Query Match             11.3%;  Score 125.6;  DB 43;  Length 810;
  Best Local Similarity   78.5%;  Pred. No. 1.6e-19;
  Matches  175;  Conservative    0;  Mismatches   45;  Indels    3;  Gaps    2;

Qy       1 atgggaccaccgaagccacgctccggatggaaaacgtggacgtgaaggaggaatggcag 60
           ||||| ||||| ||||| || ||  |||:||| || ||||||||| ||| |||||||||
Db     325 ATGGGAACCACAGAAGCTACACTAAGGATGGAAAATGTGGACGTGAGGGATGAATGGCAG 384

Qy      61 gacgaagatcttccagggccactcccagaagaaacgggggtggaactgcttggcagcccg 120
           || || ||||| ||||| || |||||||:||| || |||||||| | ||| ||  || |
Db     385 GATGAGGATCTGCCCAGACCGCTCCCAGAAGACACGGGGGTGGAGCGGCTGGGTGGCGCA 444

Qy     121 gtggaagacacatcctctcctcccaacacgctaaatttcaacggagcgcatcgtaagagg 180
           |||||||| | ||  || | ||  ||:|| || || | |||||| ||||| |||||
Db     445 NTGGAAGACTGCTC--CTACCTCCTTCACCCNGAACTTGAGCGGAGCACATCGAAAGAGA 502

Qy     181 aagacgctggtggccccagagatcaacatttnctggatcaga 223
           |||||||||||| | |||||||||||  |  ||||| || |
Db     503 AAGACGCTGGTGG-CTCAGAGATCAACATCTTCTGGACCAAA 544


Search completed: February  7, 2000, 19:32:38
Job time: 6668 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2000, 19:22:22 ; Search time 3770.17 Seconds
(without alignments)
4241.171 Million cell updates/sec

Title: US-09-174-937-4
Perfect score: 4235
Sequence: 1 gtcgacccacgcgtccgcgg ..... aaaaaaaaaagggcggccgc 4235

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	646.8	15.3	661	62	AI928493	AI928493 wp02f03 x
c	2	581.4	13.7	589	62	AI928116	AI928116 wp11c12 x
c	3	558.6	13.2	584	61	AI610536	AI610536 wb89h08 x
c	4	520	12.3	520	46	AI459935	AI459935 tk07c09 x
c	5	518	12.2	604	27	AA053602	AA053602 zk69b07 s
c	6	508.8	12.0	512	45	AI897021	AI897021 wr34h03 x
c	7	487	11.5	487	38	AA808163	AA808163 nf51f01 s

c 8	466.4	11.5	469	62	AI912605	AI912605 wo11f03.x
c 9	482	11.4	490	62	AI928170	AI928170 wo95a02.x
10	474	11.2	494	39	AA1[illegible]	AA1[illegible] z137h04.r
11	474	11.2	474	64	AL119873	AL119873 DKFZp761L
c 12	480	10.9	480	33	AA443576	AA443576 zw34h12.s
c 13	459	10.9	459	42	AI089236	AI089236 qb05c07.x
c 14	458	10.9	470	24	N24766	N24766 yx56f07.s1
c 15	454	10.7	454	40	AA927927	AA927927 om23b08.s
c 16	451.2	10.7	456	43	AI222737	AI222737 qp38d04.x
c 17	448.6	10.6	463	60	AI806606	AI806606 wf14e04.x
18	445.8	10.5	572	22	H05195	H05195 yl85f06.r1
c 19	445.4	10.5	447	42	AI093435	AI093435 qb08a12.x
c 20	444	10.5	444	64	AW073497	AW073497 xa96f04.x
c 21	425	10.0	425	40	AA987276	AA987276 os11b08.s
c 22	430	10.2	442	24	N26684	N26684 yx93f05.s1
c 23	425.6	10.0	432	38	AA757155	AA757155 ah55h01.s
c 24	417.4	9.9	445	27	AA044289	AA044289 zk51b04.s
c 25	419	9.9	427	42	AI124060	AI124060 qa99g07.s
c 26	414	9.8	414	62	AI928129	AI928129 wp11d11.x
c 27	411.2	9.7	471	29	AA131819	AA131819 zl37h04.s
28	408	9.6	419	27	AA058550	AA058550 zk69b07.r
c 29	404.2	9.5	409	46	AI457271	AI457271 tk06d06.x
c 30	399.8	9.4	403	102	AQ349510	AQ349510 RPCI11-10
31	399.2	9.4	473	29	AA147504	AA147504 zl50c06.r
c 32	397	9.4	409	39	AA779002	AA779002 ab70e02.r
c 33	395.8	9.3	456	24	H99375	H99375 yx20d02.s1
c 34	392	9.3	403	39	AA861860	AA861860 ak39d12.s
35	385.4	9.1	494	25	N45486	N45486 yy59c08.r1
c 36	379.8	9.0	400	61	AI824277	AI824277 wj08b04.x
c 37	359.8	8.5	554	22	H16385	H16385 ym24a07.s1
c 38	356.6	8.4	583	29	AA155673	AA155673 zl50c06.s
c 39	354	8.4	443	27	AA044258	AA044258 zk51f04.s
c 40	345.2	8.2	437	22	H05155	H05155 yl85f06.s1
41	345	8.1	434	22	R36574	R36574 yh39a04.r9
42	343.2	8.1	430	22	R25052	R25052 yg42a08.r1
43	331.8	7.8	340	22	AA350239	AA350239 EST57522
c 44	328.2	7.7	349	63	AI953994	AI953994 wx78c08.x
c 45	320	7.6	332	46	AI453941	AI453941 tj47f08.x

ALIGNMENTS

```
RESULT   1
AI928493/c
LOCUS       AI928493        661 bp    mRNA            EST       01-SEP-1999
DEFINITION  wp02f03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2463677 3',
            mRNA sequence.
ACCESSION   AI928493
VERSION     AI928493.1  GI:5664457
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188271.
            Contact: Robert Strausberg  Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2463677"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified p[illegible] Site; N.[illegible] Site; RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro following HAP
                     purification; this DNA was used as trimer in a subtractive
                     hybridization reaction; The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1456[illegible]-1428911, 1456007-1456775, and
                     1500552-1503[illegible]); Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      220 a    127 c    105 g    209 t
ORIGIN
```

```
  Query Match           15.3%;  Score 646.8;  DB 62;  Length 661;
  Best Local Similarity   99.5%;  Pred. No. 1.9e-120;
  Matches  659;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy   3535 agataggtagacagaggtaaacattgcagtgatttagctctctaggctcttgcgttgtgaa 3594
          ||||||||||||||||||||||| |||||||||||||||||||| | |||||||||||||||
Db    661 AGATAGGTAGACAGAGGTAAACA TGACATCATTTAGCTCTCTAGCTCTTTCGACTTTCAA 603

Qy   3595 cattcttgctttggttctgacttttgaaaatgctttgcatttctttctatatatctatta 3654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CATTCTTGCTTTGGTTCTGACTTCTCGAACTCGTTTCGATTTCTCTATATCATCTTTAA 543

Qy   3655 ttaagggaaccaagggtcattgggcaaaagcattgtttctcaaaagctcctttatttaa 3714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 TTAAGGGAACCAAGGGTCATTCGCCCAAAAGCATTGTTTCTCAAAGCTCCTTTATTTAA 483

Qy   3715 agaaagaacagaaatttgcacagaacatagtgtcaaggagtgagaaagtttgtttgaaaa 3774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 AGAAAGAACAGAAATTTGCACAGAACATAGTCTCAAGGAGTGAGAAACTTTGTTTGAGAA 423

Qy   3775 cagtagctcagtgtggaagaaaatcctgaagtttctgttgaagccatacaatgtttgtat 3834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 CAGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTTGTAT 363

Qy   3835 gggttactctctaagacattctctgaggatgtgtgaggaagtcactactcctagctttgt 3894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 GGGTTACTCTCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCTTTGT 303

Qy   3895 taagatgtaatttttaaatattcagttatggtactatgtttgcaactctcgtgttatcaca 3954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TAAGATGTAATTTTAAATATTCAGTTATGGTACTATGTTTGCAACTCTCGTGTTATCACA 243

Qy   3955 atgcctcagtagtttgttccctttaaaaaatttagatgtgcacaaattaatctttatat 4014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 ATGCCTCAGTAGTTTGTTCCCTTAAAAAATTTAGATGTGCACAAATTAATCTTTATATA 183

Qy   4015 atctaaaggttttctatcatgcattgattgctcagaataaaagtgtctgttaaattta 4074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ATCTAAAGGTTTTCTATCATGCATTGATTGCTCAGAATAAAAGTGTCTGTTAAATTTA 123

Qy   4075 ttttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacaaa 4134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TTTTGGTAAATAAATTCTCCATAATGTAGATTAATAATATAAAAGTCTTTAATGACAAA 63

Qy   4135 tatatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaa 4194
          |||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||
Db     62 TATATCTATATAGCCTCCCTGTATAATTAGAAATAAAAATAATCTCCGTTTTTTTTTT 3

Qy   4195 aa 4196
          ||
```

```
Db      2 AA 1


RESULT   2
AI928119/c
LOCUS       AI928119      589 bp    mRNA            EST       02-SEP-1999
DEFINITION  wp11c12.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2464534 3',
            mRNA sequence.
ACCESSION   AI928119
VERSION     AI928119.1  GI:5664083
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187897.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1..589
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2464534"
                     /clone_lib="NCI_CGAP_Kid12"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI.
                     Plasmid DNA from the normalized library NCI_CGAP_Kid5 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1323912-1325831, 1471368-1472903 and
                     1492104-1493255). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      199 a    101 c    106 g    183 t
ORIGIN


  Query Match          13.7%;  Score 581.4;  DB 92;  Length 589.
  Best Local Similarity   99.8%;  Pred. No. 2.8e-107;
  Matches  582;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0.

Qy   1569 ctttctaatccacgtcaagtttatttttagttgttcaaattgggcattatatttttatgtaaa 1628
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    589 CTTTCTAATCCACGTCAAGTTTATTTTTAGTTGTTCAAATTGGGCATTATATTTTTATGTAAA 530

Qy   1629 ttggtcttttaagatcatttttcctgatgaaatgttggtgaccaccacattgtgaaatttaa 1688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    529 TTGGTCTTTTAACATCATTTTCCTGATGAATGTTGGTGACCACCACATTGTGAAATTTAA 470

Qy   1689 gaatccgtgttgcatgtttggtagctctctcagtttcaggccataaactcagctccagag 1748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    469 GAATCCGTGTTGCATGTTTGGTAGCTCTCTGAGTTTCAGGCCATAAACTCAGCTCCAGAG 410

Qy   1749 gttacctttaagtgccaagaagtcaagtgccaggtgtttaatgtaaaatttttgtg 1808
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    409 GTTACCTTTTAAGTGCCAAGAAGTCAAGTGCCAAGGTGTTTACTCAAAAATTCATTTTTTA 350

Qy   1809 gcattcagttattcatgaattcctctctcgcatgcattataaaaaaatcatctttttttaaa 1868
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    349 GCATTCAGTTATTCATGAAATTCCTTTTTTGCATGCATTATAAAAAATCATCTTTTTTTAAA 290

Qy   1869 acacgtaatctgatcataggcttaaaattaaatatgagtattttttttaatgtaaaat 1928
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 ACACGTAATCTGATCATAGGCTTAAAATTAAATATGAGTATTTTTTTAATGTAAAAT 230

Qy   1929 atttcctttatagtcttcatatgccctttaaaatgtcaacaagaattttaaagtttgtaaaa 1988
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 ATTTCCTTTATAGTCTTCATATGCCCTTTAAAATGTCAACAAGAATTTTAAAGTTTGTAAAA 170

Qy   1989 ctctagtgaagtgggtgtgcaaaatacaagtaagttgtgtttgtgaatgtgaaatgaaaa 2048
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    169 CTCTAGTGAGGTGGGTGTGCAAAATACAAGTAAGTTGTGTTTGTGAATGTGAAATGAAAA 110

Qy   2049 tggttttacattttgggttgcaaaaattttttttgaatatttgatgacacatgaaaatt 2108
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    109 TGGTTTTACATTTTGGGTTGCAAAAATTTTTTTTGAATATTTGATGACACATGAAAATT 50

Qy   2109 attcaaatgttagtgccgataaataaagtggtactgaaacaca 2151
           ||||||||||||||||||||||||||||||||||||||||||||
Db     49 ATTCAAATGTCAGTGCCGATAAATAAAGTGGTACTGAAACACA 7


RESULT   3
AI810539/c
LOCUS       AI810539      584 bp    mRNA            EST       07-JUL-1999
DEFINITION  wb89h08.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2512895 3',
            mRNA sequence.
ACCESSION   AI810539
VERSION     AI810539.1  GI:5397115
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948315.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1..584
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2512895"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3 Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr28 was prepared, and ss
```

circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101142-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 197 a 100 c 104 g 183 t
ORIGIN

Query Match 12.2%; Score 572.5; DB 61; Length 584;
Best Local Similarity 99.1%; Pred. No. 1.1e-102;
Matches 572; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy    1575 aatccacgtcaacttttattttagttgtcaaattgggcattatattttatgtaaattggtc 1634
           ||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||||
Db     584 AATCCACGTCAACTTTATTTTAGTTGTCAAATTGGGCATTATATTTTATGTAAATTGGTC 525

Qy    1635 ttttaacatcattttcctgatgaaatgttggtgaccaccacattgtgaaaatttaagaatcc 1694
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     524 TTTTAACATCATTTTCCTGATGAATGTTGGTGACCACCACATTGTGAAATTTAAGAATCC 465

Qy    1695 gtgtttcatgttttgttagctctctgagtttcaaaccataaactcagctccagaggttacc 1754
           |||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     464 GTGTTGCATGTTTGGTAGCTCTCTGAGTTTCAATGCCATAAACTCAGCTCCAGAGGTTACC 405

Qy    1755 ttttaagtgccaagaactcaagtgcaagtgtgcctactcaaaaatcatttggtagcattc 1814
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     404 TTTTAAGTGCCAAGAACTCAAGTGCAAGGTGTGCTACTCAAAAATCATTTGGTAGCATTC 345

Qy    1815 agttatttgtgaattcctctctgcatgtattctaaaaagtgatctgctttaaaacaccg 1874
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     344 AGTTATTTATGAATTCCTTCTCTGCATGTATTATAAAAAGTGATCTGCTTTAAAACACCG 285

Qy    1875 taatctgatcataggcttaaaattaaatatgaatattacttttcatgtacaaaatatttcc 1934
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     284 TAATTTGATCATAGGCTTAAAATTAAATATGAATATTACTTTTCATGTACAAAATATTTCC 225

Qy    1935 tttatagtcttcatatgcccttt-aaaatgctaacaagatttcaagtctgtaggcctcta 1993
           |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db     224 TTTATAGTCTTCATATGCCCTTTAAAAATCCTAACAAGATTTCAAGTCTGTAGGCCTCTA 165

Qy    1994 gtgaggtgggtggcaaaccacagctaagtctgctcaccactgcaagctaagaatggtt 2053
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     164 GTGAGGTGGGGTGGCAAACCACAGCTAAGTCTGCTCACCACTGCAAGCTAAGAATGGTT 105

Qy    2054 tttacattttgggttgggaaaaatttttttttgaatatttcatgacacatgaaaattattca 2113
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     104 TTTACATTTTGGGTTGGAAAAATTTTTTTTTGAATATTTCATGACACATGAAAATTATTCA 45

Qy    2114 aatgttagtgccgataaataaagtggtactgaaacac 2150
           |||||||||||| |||||||||||| || |||||
Db      44 AATGTTAGTGCCATAAAATAAAGTGGTACTGAAACCC 8
```

RESULT 4
AI458305/c
LOCUS AI458305 520 bp mRNA EST 14-APR-1999
DEFINITION [illegible] NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2150220 3',
mRNA sequence.
ACCESSION AI458305
VERSION AI458305.1 GI:4310884
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 520)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 19 1999 this sequence version replaced gi:3139779.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.; Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 641 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450
FEATURES Location/Qualifiers
source 1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2150220"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417999, and 1722924-1722999). Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 177 a 100 c 78 g 165 t
ORIGIN

Query Match 12.3%; Score 580; DB 46; Length 520;
Best Local Similarity 100.0%; Pred. No. 6.4e-95;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    3664 ccaaggggtcattgggggcaaaagcaatgttctcaaagcaaattggatgaaaaaaac 3723
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 CCAAGGGGTCATTGGGGCAAAACCATGTTTTTCAAAGCCCTTGATGAAAAAAAC 461

Qy    3724 agaaatttgcacagaagatagtgtgcaaaagtgaagaaaagttatgtgaaaaaaaaaat 3783
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 AGAAATTTGCACAGAAGATAGTGTCAAACCAGTCAGAAAGTTTGTTTCATCCTAATAATC 401

Qy    3784 agtgtggaagaaaatcctgaagttctgttgaagccatacaatgttctatgggattactc 3843
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     400 AGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTCTATGGGATTACTC 341

Qy    3844 tctaagacattctctgaggtgtgtgaggaaagtcactactcctagccttttgttaaatatca 3903
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     340 TCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCCTTTTGTTAAATATCA 281

Qy    3904 attttaaatattcagttatggtactatgtttgcaaactctcgtcttatgacaatagcctgaa 3963
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 ATTTTAAATATTCAGTTATGGTACTATGTTTGCAAACTCTCGTCTTATGACAATAGCCTGAA 221

Qy    3964 tagtttgttccgttagaaacatttagaatgtgcacaaattaatctttttatatat tgaaat 4023
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     220 TAGTTTGTTCCGTTAGAAACATTTAGATGTGCACAAATTAATCTTTTTATATAT TAAAAT 161

Qy    4024 ttttctatcatgcattggattgctaaaataaagtgtctgttagacttcgtttttgtaa 4083
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 TTTTTCTATCATGCATTGGATTGCTACAATAAAGTGTCTGTTAGACTTCGTTTTTGTAA 101

Qy    4084 ataaattctccataatgtagattaatatataaaagtctttaatgacacaatatatgtat 4143
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100 ATAAATTCTCCATAATGTAGATTAATATATAAAAGTCTTTAATGACACAATATATGTAT 41

Qy    4144 atagcctcactgtataattcagaaataaaattgattcta 4183
```

```
Db     40 ATAGCCTCACTGTATAAITCAGAAATAAAAATTGATTCTG 1


RESULT   5
AA053302/c
LOCUS       AA053302      604 bp    mRNA            EST       11-MAY-1997
DEFINITION  zb69b07.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:488053 3', mRNA sequence.
ACCESSION   AA053302
VERSION     AA053302.1  GI:1545761
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 987   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 349.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Homo sapiens"
                     /db_xref="GDB:3802640"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:488053"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCTCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      191 a    131 c    103 g    169 t     10 others
ORIGIN


  Query Match             12.2%;  Score 513,  DB 27,  Length 604;
  Best Local Similarity   93.7%;  Pred. No. 1.6e-94;
  Matches  567;  Conservative    0;  Mismatches   34;  Indels    4;  Gaps    3;

Qy   3582 cttgcactctgaacattcttgctttggttctgacttctgggaactgctttgcatttctcc 3641
          ||  |||||| |||||| ||    | ||||| ||||| | |||| ||||| | ||
Db    603 CIGCCACICNGAACAINCI--GCTGGGTTCGNACTTCGGGGANCGGCTTGGCATTCCCCC 546

Qy   3642 tatagatctgtagttaagggaaccaa-ggggtcattggggcaaaagcattgtttctcaaa 3700
          |||||||.|.||||   |||| || | ||||  || ||| |||| ||||||||||| |||
Db    545 TATAGATCIGTAGTAGGGGGAGCCCNAGGGGGTCATGGGGNCAAAGGCATTGTTTCCCAAG 486

Qy   3701 gctccttgattaagagaaaga-acaaaaatttgcatgaaaaatgtatgtctaaaaaatgaaa 3759
          || |||| |||| | ||| || | |||| | |||| |||| | |||| |||| | |||| |
Db    485 GCNCCTGGATTAGGNGAAGGACCCNAAATTTTCATAAAAATGTATGTCTAAAAAATNAAA 426

Qy   3760 aagtttgtttgagggcagtagcttcatgtgaaaaaaaatgtcatgaaaaaaaaaaaaaaaa 3819
          |||||||||||| |||||||||| | | | | | | | | | | | | | | | | | |
Db    425 AAGTTTGTTTGAGGGCAGTAGCTCATGTTCGAAAAAAATGTCATGAAAAAAAAAAAAAAA 366

Qy   3820 atacaatgttctatgggggttactctctaagacattttttgaaaatgtgtgaaaaaatgtct 3879
          |||||||||||.|||||||||  | | | | | | | | | | | | | | | | | | |
Db    365 ATACAAIGTTCTATGGGGTTAC.CTCTAAGA.ATTTTTTGAAA.TGTGTGAAAAATGTCT 306

Qy   3880 actcctagcctttgttaacatgtaatttttaaatattcagttatatatactatttttgcaac 3939
          | | | ||||.|||| |||| | | | | | | | | | | | | | | | | | | |
Db    305 ACTCCTAGCCTTTGTTAACATGTAATTTTTAAATATTCAGTTATATATACTATTTTTGAAC 246

Qy   3940 tctcgtcttatcacaatgcctcaatattttgttcctttaaaaaatttaatatgtgtgcaaa 3999
          |||| |||||.|||| |||||  | | | | | | | | | | | | | | | | | | |
Db    245 TCTCGTCTTATCACAATGCCTCAATATTTTGTTCCTTTAGAAACATTTAATATGTGTACAA 186

Qy   4000 attaatcttttatatatctaaaggttttctatcatgcattgaattgctcagaaaaaaagt 4059
          |||||||||||||||||||||  | | | | | | | | | | | | | | | | | | |
Db    185 ATTAATCTTTTATATATCTAAAGTTTTCTATCATGCATTGAATTGCTCAGAAATAAAGT 126

Qy   4060 gtctgttagacttcgttttggtaaataaattctccataatgtagattaaatgatatgaaaaa 4119
          |||||||||||||||||||||  | | | | | | | | | | | | | | | | | | |
Db    125 GTCTGTTAGACTTCGTTTTGGTAAATAAATTTCTCCATAATGTAGATTAAATGATATGAAAAA 66

Qy   4120 tctttaatgacacaatatatctatatagcttgatgtatgattgtgcaatgaaaattcat 4179
          |||||||||||||||||||||  | | | | | | | | | | | | | | | | | | |
Db     65 TCTTTAATGACACAATATATCTATATAGCTTGCTTTATGATTGTGCAATGAAAATTCAT 6

Qy   4180 tctgc 4184
          |||||
Db      5 TCTGC 1


RESULT   6
AI337921/c
LOCUS       AI337921      512 bp    mRNA            EST       13-FEB-1999
DEFINITION  qt34h03.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1949909 3', mRNA sequence.
ACCESSION   AI337921
VERSION     AI337921.1  GI:4074848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 723   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1949909"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
```

```
                oligo(dT) primer [5'
                AACTGGAAGAATTCGCGGCCGCTTTTTTTTTTTTTTTTTT 3'],
                double stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                went through one round of normalization. Library
                constructed by M. Fatima Bonaldo."
BASE COUNT      174 a     92 c     77 g    169 t
ORIGIN

  Query Match            12.0%;  Score 518.8;  DB 45;  Length 512;
  Best Local Similarity  99.6%;  Pred. No. 1.2e-92;
  Matches  510;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   3683 aaagcattgtttctcaaagctccttgattaagagaaagaacagaaatttgcacagaagat 3742
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    512 AAAGCATTGTTTCTCAAAGCTCCTTGATTAAGAGAAAGAACAGAAATTTGCACAGAAGAT 453

Qy   3743 agtgtcaaggagtgagaaagtttgtttgaggcagtagctcagtgtggaagaaaatcctg 3802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    452 AGTGTCAAGGAGTGAGAAAGTTTGTTTGAGGCAGTAGCTCAGTGTGGAAGAAAATCCTG 393

Qy   3803 aagtttctgttgaagccatacaatgttctatggggttactctctaagacattctctgagg 3862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    392 AAGTTTCTGTTGAAGCCATACAATGTTCTATGGGGTTACTCTCTAAGACATTCTCTGAGG 333

Qy   3863 tgtgtgaggaagtcactactcctagcctttgttaagatgtaattttaaatattcagttat 3922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 TGTGTGAGGAAGTCACTACTCCTAGCCTTTGTTAAGATGTAATTTTAAATATTCAGTTAT 273

Qy   3923 ggtactatgtttgcaactctcgtcttatcacaatgcctcagtagtttgttccccttagaaa 3982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 GGTACTATGTTTGCAACTCTCGTCTTATCACAATGCCTCAGTAGTTTGTTCCCTTAGAAA 213

Qy   3983 catttagatgtgcacaaattaatcttttatatatctaaaggtttttctatcatgcattgg 4042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 CATTTAGATGTGCACAAATTAATCTTTTATATATCTAAAGGTTTTTCTATCATGCATTGG 153

Qy   4043 attgctcagaataaagtgtctgttagacttcgttttggtaaataaattctccataatgta 4102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 ATTGCTCAGAATAAAGTGTCTGTTAGACTTCGTTTTGGTAAATAAATTCTCCATAATGTA 93

Qy   4103 gattaataatataaaagtctttaatgacacaatatatctatatagcctcactgtataatt 4162
          |||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db     92 GATTAATAATATAAAAGTCTTTAATGACCCAATATATCTATATAGCCTCCCTGTATAATT 33

Qy   4163 cagaaataaaaattgattctgcaaaaaaaaaa 4194
          ||||||||||||||||||||||||||||||||
Db     32 CAGAAATAAAAATTGATTCTGCAAAAAAAAAA 1

RESULT  7
AA808163/c
LOCUS       AA808163      487 bp    mRNA            EST       19-FEB-1998
DEFINITION  of51f01.s1 NCI_CGAP_CNS1 Homo sapiens cDNA clone IMAGE:1427737 3',
            mRNA sequence.
ACCESSION   AA808163
VERSION     AA808163.1  GI:2877569
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152329.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            unknown library type
            Insert Length: 858   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1..487
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1427737"
                     /clone_lib="NCI_CGAP_CNS1"
                     /tissue_type="substantia nigra"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT4; Site_1: SalI;
                     Site_2: NotI. Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.0 kb."
BASE COUNT      165 a     88 c     69 g    165 t
ORIGIN

  Query Match            11.5%;  Score 487;  DB 28;  Length 487;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-88;
  Matches  487;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3710 ttaagagaaagaacagaaatttgcacagaagatagtgtcaaggagtgagaaagtttgttt 3769
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 TTAAGAGAAAGAACAGAAATTTGCACAGAAGATAGTGTCAAGGAGTGAGAAAGTTTGTTT 428

Qy   3770 gaggcagtagctcagtgtggaagaaaatcctgaagtttctgttgaagccatacaatgttc 3829
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 GAGGCAGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTC 368

Qy   3830 ctatggggttactctctaagacattctctgaggtgtgtgaggaagtcactactcctagcc 3889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 CTATGGGGTTACTCTCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCC 308

Qy   3890 tttgttaagatgtaattttaaatattcagttatggtactatgtttgcaactctcgtctta 3949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 TTTGTTAAGATGTAATTTTAAATATTCAGTTATGGTACTATGTTTGCAACTCTCGTCTTA 248

Qy   3950 tcacaatgcctcagtagtttgttcccttagaaacatttagatgtgcacaaattaatcttt 4009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 TCACAATGCCTCAGTAGTTTGTTCCCTTAGAAACATTTAGATGTGCACAAATTAATCTTT 188

Qy   4010 tatatatctaaaggtttttctatcatgcattggattgctcagaataaagtgtctgttaga 4069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 TATATATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTAGA 128

Qy   4070 cttcgttttggtaaataaattctccataatgtagattaataatataaaagtctttaatga 4129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 CTTCGTTTTGGTAAATAAATTCTCCATAATGTAGATTAATAATATAAAAGTCTTTAATGA 68

Qy   4130 cacaatatatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaa 4189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 CACAATATATCTATATAGCCTCACTGTATAATTCAGAAATAAAAATTGATTCTGCAAAAA 8

Qy   4190 aaaaaaa 4196
          |||||||
Db      7 AAAAAAA 1

RESULT  8
AI912605/c
LOCUS       AI912605      489 bp    mRNA            EST       [illegible]
DEFINITION  we11f03.x1 NCI_CGAP_[illegible] Homo sapiens cDNA clone [illegible] 3',
            mRNA sequence.
ACCESSION   AI912605
VERSION     AI912605.1  GI:5632460
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 489)
```

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1996 this sequence version replaced gi:1132602.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1..489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:234079"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      166 a     90 c     73 g    156 t      2 others
ORIGIN


  Query Match             11.5%;  Score 488.4;  DB 62;  Length 489;
  Best Local Similarity   99.4%;  Pred. No. 6e-88;
  Matches  486;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    3695 ctcaaagctcctttgattaagagaaagaacagaaatttgcacagaagatagtgtcaaggag 3754
           ||| |||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db     489 CTCANAGCTCCTTGATTAAGAGAAAGAACAGAAATTTGCACAGAAGATAGTGTCAAGGAG 430

Qy    3755 tgagaaagtttgtttgagggcagtagctcactgtggaagaaaatcctgaagtttctgttg 3814
           ||| |||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db     429 TGAGAAAGTTTGTTTGAGGGCAGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTG 370

Qy    3815 aaggcatacaaatgttctatgggttactgtctaagacattcctgaggtgtgtgaggaag 3874
           ||| |||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db     369 AAGCCATACAAIGTTCTATGGGGTTACTGTCTAAGACATTCTCTGAGGTGTGTGAGGAAG 310

Qy    3875 tcactactcctagcctttgttaagatgtaatttaaatattcagttatggtactatgttt 3934
           |||||||| |||||||||||||||||| | ||||||||||||||||||||||||||
Db     309 TCACTACTCCTAGCCTTTGTTAAGATGTAATTTAAATATTCAGTTATGGTACTATGTTT 250

Qy    3935 gcaaactctcgtcttatcacaatgcctcagtaattttgttcccttagaaacatttagatgtg 3994
           ||| | |||||| |||||||| | |||||||||||||||||||||||||||||||||||
Db     249 GCAACTCTCGTCTTATCACAATGCCTCAGTAATTTTGTTCCCTTAGAAACATTTAGATGTG 190

Qy    3995 cacaaattaatcttttatatatctaaaggttttctatcatgcattggattgctcagaat 4054
           ||| |||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db     189 CACAAATTAATCTTTTATATATCTAAAGGTTTTCTATCATGCATTGGATTGCTCAGAAT 130

Qy    4055 aaagtgtctgttagacttcgttttggtaaatgaattctccataatgtagattaataatat 4114
           ||| |||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||
Db     129 AAAGTGTCTGTTAGACTTCGTTTTGGTAAATAAATTCTCCATAATGTAGATTAATAATAT 70

Qy    4115 aaaagtctttaatgacacaatatatctatattcctcactgtataattcagaaataaaaa 4174
```

```
Db      69 AAAAGTCTTTAATGACACAATA                                    10

Qy    4175 ttgattctg 4183

Db       9 TAGATTCTG 1


RESULT   9
AI928170/c
LOCUS       AI928170        490 bp    mRNA            EST     12-SEP-1999
DEFINITION  wj95a02.x1 NCI_CGAP_K1111 Homo sapiens cDNA clone IMAGE:2473051 3',
            mRNA sequence.
ACCESSION   AI928170
VERSION     AI928170.1  GI:5664134
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1999 this sequence version replaced gi:1132602.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1..490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2473051"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3-Pac (Pharmacia) with
                     a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1422376-1423921, 1456[illegible], and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      167 a     89 c     73 g    161 t
ORIGIN


  Query Match             11.4%;  Score 484.4;  DB 62;  Length 490;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-87;
  Matches  485;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    3706 ttgattaagagaaagaacagaaatttgcacagaagatagtgtcaaggagtgagaaagtt 3765
           |||| |||||||||||||||||||||||||||||| | ||||||||||||||||||||||
Db     490 TTGATTAAGAGAAAGAACAGAAATTTGCACAGAAGATAGTGTCAAGGAGTGAGAAAGTTT 431

Qy    3766 gtttgagggcagtagctcagtgtggaaagaaaatcctgaagtttctgttgaaggcatacaa 3825
           |||| |||||||||||||||||||||||||||||| | ||||||||||||||||||||||
Db     430 GTTTGAGGGCAGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAA 371
```

```
Qy   3826 tatttctatgggggttactctctaagacattctcgaggtgtgtgaggaagtcactactcct 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 TGTTCTATGGGGTTACTCTCTAAGACATTCTCGAGGTGTGTGAGGAAGTCACTACTCTT 311

Qy   3886 agccttttattaagatgtaatttttaaatattcatttatggtactatgtttgcaactctcgt 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AGCCTTTGTTAAGATGTAATTTTAAATATTCATTTATGGTACTATGTTTGCAACTCTCGT 251

Qy   3946 cttatcacaatgcctcagtagtttgttcctttagaaacatttagatgtgcacaaattaat 4005
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    250 CTTATCACAATGCGTCAGTAGTTTGTTCCTTTAGAAACATTTAGATGTGCACAAATTAAT 191

Qy   4006 cttttatatatctaaaggtttttctatcatgcattggattgctcagaataaagtgtctgt 4065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 CTTTTATATATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAGAATAAAGTGTCTGT 131

Qy   4066 tagacttcgttttggtaaataaattctccataatgtagattaatatatataaagtctttta 4125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 TAGACTTCGTTTTGGTAAATAAATTCTCCATAATGTAGATTATATATATAAAGTCTTTA 71

Qy   4126 atgacacaatatatctatatagcctcactgtataattcagaaataaaaattgattctgca 4185
          ||||||||||||||||||||| |||| |||||  |||||||||||||||||| ||||||||||
Db     70 ATGACACAATATATCTATATAGCCTCCCTGTATAATTCAGAAATAAAAAATGATTCTGCA 11

Qy   4186 aaaaaaaaaa 4195
          ||||||||||
Db     10 AAAAAAAAAA 1


RESULT  10
AA132002
LOCUS       AA132002      492 bp    mRNA            EST       14-MAY-1997
DEFINITION  zl37h04.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:504151 5', mRNA sequence.
ACCESSION   AA132002
VERSION     AA132002.1  GI:1603493
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1148  Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Homo sapiens"
                     /db_xref="GDB:3809134"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:504151"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3 Pac I Site_1; Not I
                     Site_2; Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCTTTTTTTTTTTTTTT-3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector; library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      143 a    100 c     82 g    157 t
ORIGIN


  Query Match           11.2%;  Score 474;  DB 29;  Length 482;
  Best Local Similarity  99.0%;  Pred. No. 1.2e-85;
  Matches  477;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    0;

Qy   2993 cttttgaaagcccaaattacacctctcccagaacacagtgttgagctaaatatatttatt 3052
          | |||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CATTTTGAAAGCCAAATTACACCTCTCCCAGAACACAGTGTTGAGCTAAATATATTTATT 60

Qy   3053 caatattcctgtttgttcacctatttactactttcactcagtagcatcccatttgtaaa 3112
          ||||||||||||||||||||||||| |||| |||||||| || | ||||||||||||||
Db     61 CAATATTCCTGTTTGTTCACCTATTTGCTACTTTCACTCAGTAGCATCCCATTTGTAAA 120

Qy   3113 atgaattccatggtcaccctgtcacaggaagtaatgaaaaaatccaatgtttcatgtatt 3172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATGAATTCCATGGTCACCCTGTCACAGGAAGTAATGAAAAATCCAGTGTTCATGTATT 180

Qy   3173 gtgcaaacctgagggcatagagctgttcatagagggctattgttatagccccccgcgcc 3232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGCAAACCTGAGGGCATAGAGCTGTTCATAGAGGGCTATTGTTATAGCCCCCGCGCC 240

Qy   3233 agcaacaatctcaccattatatatatattttaacttatcaggctcatctgtgaaaa 3292
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    241 AGCAACAATCTCACCATTATATATATATTTTAACTTATCAGGCTCATCTGTGAAAA 300

Qy   3293 tactcaggtggtatgctgtttgaagcctcatcttcctacatgaaaattatgtgcatttat 3352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TACTCAGGTGGTATGCTGTTTGAAGCCTCATCTTCCTACATGAAAATTATGTGCATTTAT 360

Qy   3353 cccaatgatttttgtttcagctgttctgtaggctgcataaccactctgtatatttaagtat 3412
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    361 CCCAATGATTTTTGTTTCAGCTGTTCTGTAGGCTGCATAACCACTCTGTATATTTAAGTAT 420

Qy   3413 tgctatttttattatcttaaaagacaaattaatttaattgcatgtgctagggaaaaggctat 3472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGCTATTTTTATTATCTTAAAAGACAAATTAATTTAATTGCATGTGCTAGGGAAAAGCTAT 480

Qy   3473 ca 3474
          ||
Db    481 CA 482


RESULT  11
AL119873
LOCUS       AL119873      474 bp    mRNA            EST       23-SEP-1999
DEFINITION  DKFZp761L0924_r1 761 (synonym: hary2) H. sapiens cDNA clone
            DKFZp761L0924 5', mRNA sequence.
ACCESSION   AL119873
VERSION     AL119873.1  GI:5925772
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Rasseenhoeve,J. and
            Wiemann,S.
  TITLE     EST (Bloecker, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:1417247.
            Contact: Bloecker H
```

```
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by GBF within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone is available at the RZPD in Berlin
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES              Location/Qualifiers
     source           1. .474
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DKFZp761L0924"
                      /clone_lib="761 (synonym: hamy2)"
                      /tissue_type="amygdala"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      135 a      93 c      78 g     168 t
ORIGIN


  Query Match             11.2%;  Score 474;  DB 64;  Length 474;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-85;
  Matches  474;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1484 acaaggtgtaaaaccatcatcagctaagtgcccgttttgttcttgtttaccagaatctcc 1543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACAAGGTGTAAAACCATCATCAGCTAAGTGCCCGTTTTGTTCTTGTTTACCAGAATCTCC 60

Qy   1544 tttaacttctcaaagggaagccgggcttttctaatccacgtcaacttttattttagttgtca 1603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTAACTTCTCAAAGGGAAGCCGGGCTTTTCTAATCCACGTCAACTTTATTTTAGTTGTCA 120

Qy   1604 aattgggcattatattttatgtaaattggtctttaacatcattttcctgatgaatgttg 1663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATTGGGCATTATATTTTATGTAAATTGGTCTTTAACATCATTTTCCTGATGAATGTTG 180

Qy   1664 gtgaccaccacattgtgaaatttaagaatccgtgttgcatgtttggtagctctctgagtt 1723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGACCACCACATTGTGAAATTTAAGAATCCGTGTTGCATGTTTGGTAGCTCTCTGAGTT 240

Qy   1724 tcaggccataaactcagctccagaggttacgttttaagtgccaagaactcaagtgcaaag 1783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCAGGCCATAAACTCAGCTCCAGAGGTTACGTTTTAAGTGCCAAGAACTCAAGTGCAAGG 300

Qy   1784 tggcctactcaaaaatcatttggtagcattcagttattcatgaattcctctctcgcatgc 1843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGCCTACTCAAAAATCATTTGGTAGCATTCAGTTATTCATGAATTCCTCTCTCGCATGC 360

Qy   1844 attataaaaagtgatctgctttaaaacaccgtaatctgatcataggcttaaaattaaata 1903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATTATAAAAAGTGATCTGCTTTAAAACACCGTAATCTGATCATAGGCTTAAAATTAAATA 420

Qy   1904 tgagtattactttcatgtacaaaatatttcctttatagtcttcatatgccctttt 1957
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGAGTATTACTTTCATGTACAAAATATTTCCTTTATAGTCTTCATATGCCCTTT 474


RESULT  12
AA443576/c
LOCUS       AA443576      460 bp    mRNA            EST       03-JUN-1997
DEFINITION  zw34h12.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:771239 3', mRNA sequence.
ACCESSION   AA443576
VERSION     AA443576.1  GI:2156251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,J., Just,S.,
            Kucaba,T., Lacy,M., Le,N., Le Mon,D., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptos,M., Tan,F., Theising,B.,
            White,T., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On May 8, 1995 this sequence version replaced gi:881972.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. BT from Amersham
            High quality sequence stop: 416.
FEATURES              Location/Qualifiers
     source           1. .460
                      /organism="Homo sapiens"
                      /db_xref="GDB:5981115"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:771239"
                      /clone_lib="Soares ovary tumor NbHOT"
                      /sex="Female"
                      /tissue_type="ovarian tumor"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: ovary; Vector: pCMV-S (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer (5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTT 3'),
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pCMV vector
                      (Pharmacia). Library constructed by Bento Soares and
                      M.Fatima Bonaldo. "
BASE COUNT      163 a      85 c      67 g     145 t
ORIGIN


  Query Match             10.9%;  Score 461;  DB 64;  Length 474;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-83;
  Matches  460;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3724 agaaatttgcacagaagatagtgtcaaccagttagaaagtttcttttgaggcccatgctc 3783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 AGAAATTTGCACAGAAGATAGTGTCAAGCAGTTAGAAAGTTTCTTTTGAGGCCATGCTC 401

Qy   3784 agtgtggaagaaaatcctgaagtttctcttgaagcgatacaatgttctatagggttgatc 3843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 AGTGTGGAAGAAAATCCTGAAGTTTCTCTTGAAGCGATACAATGTTCTATAGGGTTGATC 341

Qy   3844 tctaagacattctctgaggtgtgtgaggaagtcactactctagcctttgttaggatgta 3903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 CCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCCTTTGTTAACATGTA 281

Qy   3904 attttaaatattcagttatggtactatctttgcaaatctccatctatcacaatgcctcaa 3963
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 ATTTTAAATATTCAGTTATGGTACTATCTTTGCAAATCTCCATCTATCACAATGCCTCAA 221

Qy   3964 tagtttgttccctagaaacattagatgtgcacaaattaatcttttatatatctaaaat 4023
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 TAGTTTGTTCCCTTAGAAACATTAGATGTGCACAATTAATCTTTTTATATATTAAAAG 161

Qy   4024 ttttctatcatgcattggattctcagaatcaagtgtctgttgaatcctgtttggtaa 4083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 TTTTTCTATCATGCATTGGATTGCTCAGAATCAAAGTGTCTGTTGAGACTTGGTTTGGTAA 101

Qy   4084 ataaattctccataatgtagattaatcatatcaaaatctctttaatgacacaatatatctat 4143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100 ATAAATTCTCCATAATGTAGATTAATCATATCAAAATCTCTTTAATGACACAATATATCTAT 41
```

```
Qy   4144 atagcctcactgtataattcagaaaataaaaattgattctg 4183
          ||||||||||||||||||||||||||||||||||||||||||
Db     40 ATAGCCTCACTGTATAATTCAGAAATAAAAATTGATTCTG 1


RESULT  13
AI089236/c
LOCUS       AI089236       459 bp    mRNA            EST       05-OCT-1998
DEFINITION  qb05e07.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1695372 3', mRNA sequence.
ACCESSION   AI089236
VERSION     AI089236.1  GI:3428295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:802957.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1000   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 442.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1695372"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer (5'
                     AACTGGAAGAATTGCGGCCGCTTTTTTTTTTTTTTTT 3'),
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      160 a     83 c     67 g    149 t
ORIGIN


  Query Match            10.8%;  Score 459;  DB 42;  Length 459;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-82;
  Matches  459;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3735 cagaagatagtgtcaagggagtgagaaagtttgtttgaggcagtagctcagtgtggaaga 3794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CAGAAGATAGTGTCAAGGAGTGAGAAAGTTTGTTGAGGGCAGTAGCTCAGTGTGGAAGA 400

Qy   3795 aaatcctaaagtttctattgaagccatacaattttctatgggattactctctaagacatt 3854
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 AAATCCTGAAGTTTCTGTTGAAGCCATACAATTTTCTATGGGATTACTCTCTAAGACATT 340

Qy   3855 ctctgaagtgtgtgaggaagtcactactcctacctttgttaagatgtaattttaaatat 3914
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 CTCTGAGGTGTGTGAGGAAGTCACTACTCCTACCTTTGTTAAGATGTAATTTTAAATAT 280

Qy   3915 tcagttatggtactatgttttgcaaatctctgtctttatgacaatgcctcagtagttttgttcc 3974
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 TCACTTATGGTACTATGTTTTGCAAATCTCTGTCTTTATGACAATGCCTCAGTAGTTTTGTTCC 220

Qy   3975 cttagaaacatttagatgtgcacaaattaatctttatatatctaaaggttttctattca 4034
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 CTTAGAAACATTTAGATGTGCACAAATTAATCTTTTATATATCTAAAGGTTTTCTATTCA 160

Qy   4035 tgcattggattgctcagaataaagtgtctgttagacttcgttttggtaaataaatttctgc 4094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 TGCATTGGATTGCTCAGAATAAAGTGTCTGTTAGACTTCGTTTTGGTAAATAAATTTCTGC 100

Qy   4095 ataatgtagattaataatatataaaagtctttaatgacacaatatatctatataactgact 4154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 ATAATGTAGATTAATAATATAAAAGTCTTTAATGACACAATATATCTATATAACTGACT 40

Qy   4155 gtataattcagaaataaaaattgattctgcaaaaaaaaa 4193
          |||||||||||||||||||||||||||||||||||||||
Db     39 GTATAATTCAGAAATAAAAATTGATTCTGCAAAAAAAAA 1


RESULT  14
N24766/c
LOCUS       N24766         470 bp    mRNA            EST       28-DEC-1995
DEFINITION  yx56f07.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone
            IMAGE:265765 3', mRNA sequence.
ACCESSION   N24766
VERSION     N24766.1  GI:1138916
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785245.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 432
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1545   Std Error: 0.00
            Seq primer: m13 -40 forward
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /db_xref="GDB:3875407"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:265765"
                     /clone_lib="Soares melanocyte 2NbHM"
                     /sex="Male"
                     /tissue_type="melanocyte"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTT 3'),
                     double-stranded cDNA was size selected (>1.0 kb), Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. RNA from normal foreskin melanocytes
                     (FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT      164 a     87 c     69 g    150 t
```

ORIGIN

Query Match 10.8%; Score 458; DB 24; Length 470;
Best Local Similarity 99.8%; Pred. No. 1e-82;
Matches 408; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 3775 gaaagaacagaaatttgcacagaagatagtgtcaaggagtgagaaagtttgtttgagggc 3834
Db 470 GAAAGAACAGAAATTTGCACAGAAGATAGTGTCAAGGAGTGAGAAAGTTTGTTTGAGGGC 411

Qy 3835 agtagctcagtgtggaagaaaatcctgaagtttctgttgaagccatacaatgttc-tatg 3893
Db 410 AGTAGCTCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTCTTATG 351

Qy 3894 gggttactctctaagacattctctgaggtgtgtgaggaagtcactactcctagccctttgt 3953
Db 350 GGGTTACTCTCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCCTTTGT 291

Qy 3954 taagatgtaatttttaaatattcagttatggtactatgtttgcaactctcgtcttatcaca 4013
Db 290 TAAGATGTAATTTTAAATATTCAGTTATGGTACTATGTTTGCAACTCTCGTCTTATCACA 231

Qy 4014 atgcctcagtagtttgttcccttagaaacatttagatgtgcacaaattaatctttttatat 4073
Db 230 ATGCCTCAGTAGTTTGTTCCCTTAGAAACATTTAGATGTGCACAAATTAATCTTTTATAT 171

Qy 4074 atctaaaggtttttctatcatgcattggattgctcagaataaagtgtctgttagacttcg 4133
Db 170 ATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTAGACTTCG 111

Qy 4134 ttttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacacaa 4193
Db 110 TTTTGGTAAATAAATTCTCCATAATGTAGATTAATAATATAAAAGTCTTTAATGACACAA 51

Qy 4194 tatatctatatagcctcactgtataattcagaaataaaaattgattctgc 4243
Db 50 TATATCTATATAGCCTCACTGTATAATTCAGAAATAAAAATTGATTCTGC 1

RESULT 15
AA927927/c
LOCUS AA927927 454 bp mRNA EST 10-JUN-1998
DEFINITION em:zb09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:154184? 3', mRNA sequence.
ACCESSION AA927927
VERSION AA927927.1 GI:3076671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045341.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1080 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 429.
FEATURES Location/Qualifiers
source 1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:154184?"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal [illegible] NbHL19W, testis NHT, and brain NOT_CGAP_GC5); cells mixed, and [illegible] made in [illegible]; FSE: [illegible] purification; this EST was used as tracer in a subtractive hybridization [illegible]. The driver was PCR-amplified cDNAs from pools of 5'... clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297486-3 2 [illegible], 726202-726[illegible], 726408-728711, and 724094-[illegible]. Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 160 a 84 c 68 g 142 t
ORIGIN

Query Match 10.7%; Score 454; DB 41; Length 454;
Best Local Similarity 100.0%; Pred. No. 1.6e-81;
Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3732 gcacagaaagataagtgtcaaggagtgagaaagtttgtttgagggcagtagctcagtgtg 3791
Db 454 GCACAGAAGATAGTGTCAAGGAGTGAGAAAGTTTGTTTGAGGGCAGTAGCTCAGTGTG 395

Qy 3792 agaaaatcctgaagtttctgttgaagccatacaatgttctatgggttactctctaagac 3851
Db 394 AGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTCTATGGGTTACTCTCTAAGAC 335

Qy 3852 attctctgaggtgtgtgaggaagtcactactcctagccctttgttaagatgtaatttttaa 3911
Db 334 ATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCCTTTGTTAAGATGTAATTTTTAAA 275

Qy 3912 tattcagttatggtactatgtttgcaactctcgtcttatcacaatgcctcagtagtttgt 3971
Db 274 TATTCAGTTATGGTACTATGTTTGCAACTCTCGTCTTATCACATGCCTCAGTAGTTTGT 215

Qy 3972 tcccttagaaacatttagatgtgcacaaattaatctttttatatatctaaaggttttttcta 4031
Db 214 TCCCTTAGAAACATTTAGATGTGCACAAATTAATCTTTTATATATCTAAAGGTTTTTCTA 155

Qy 4032 tcatgcattggattgctcagaataaagtgtctgttagacttcgttttggtaaataaatttct 4091
Db 154 TCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTAGACTTCGTTTTGGTAAATAAATTTCT 95

Qy 4092 ccataatgtagattaataatatataaaagtctttaatgacacaatatatctatatagcctc 4151
Db 94 TCCATAATGTAGATTAATAATATATAAAAGTCTTTAATGACACAATATATCTATATAGCCTC 35

Qy 4152 actgtataattcagaaataaaaattgattctgca 4185
Db 34 ACTGTATAATTCAGAAATAAAAATTGATTCTGCA 1

Search completed: February 7, 2000, 19:39:05
Job time: 6695 sec

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 20:51:01 ; Search time 284.47 Seconds
(without alignments)
332.452 Million cell updates/sec

Title: US-09-174-937-9
Perfect score: 378
Sequence: 1 atgctgaagagctgtagtcg.....taatttggcaaggccaaagt 378

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_99.*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266.4	70.5	1227	1	X39873	Gastric cancer ass
2	266.4	70.5	816	1	X39876	Gastric cancer ass
3	34.2	9.0	1700	1	X29771	L lactis HsdM gene
4	34	9.0	1386	1	T38228	Apoptosis-regulati
5	34	9.0	1386	1	X04120	Human Nip2 encodin
6	31	8.2	3001	1	T61583	Malolactic enzyme
c 7	31	8.2	824	1	V31262	E. coli O96 pathog
c 8	30.4	8.0	1610	1	V90545	Nucleotide sequenc
9	30.2	8.0	3658	1	X13324	Enterococcus faeca
c 10	30.2	8.0	3990	1	V72213	Human NT2-1A DNA.
c 11	30.2	8.0	4417	1	V72214	Human NT2-2A DNA.
12	30	7.9	4383	1	Q10263	pZ130 contg. Calge
13	30	7.9	4383	1	Q10319	Calgene lambda 140
14	30	7.9	4383	1	Q35143	Calgene lambda 140
15	30	7.9	4383	1	Q34941	Calgene Lambda 140
16	30	7.9	3528	1	I31824	Tomato pZ130 genom
17	30	7.9	3528	1	I48616	Calgene Lambda 140
18	29.8	7.9	669	1	Q03334	Fragment of post-t
19	29.8	7.9	569	1	Q42652	PT-NANB virus lamb
20	29.8	7.9	1759	1	T71019	Type I, p80 IL-1-r
21	29.8	7.9	1100000	1	X20248_09	Continuation (9 of
c 22	29.2	7.7	1100000	1	V21209_11	Continuation (15 o
23	29.2	7.7	1505	1	V59987	DNA encoding Staph
24	29.2	7.7	1076	1	V74352	Staphylococcus aur
25	29	7.7	2627	1	V52087	Helicobacter polyp
26	29	7.7	11427	1	X13102	Enterococcus faeca
27	29	7.7	2627	1	X14421	H. pylori GHP1 615
c 28	28.8	7.6	1373	1	V18600	Homo sapiens tumou
c 29	28.8	7.6	1036	1	V47623	TNF related endoth
c 30	28.8	7.6	19290	1	V52219	Streptococcus pneu
31	28.8	7.6	924	1	V71744	Human V3 loop HIV
c 32	28.6	7.6	5108	1	Q53403	Sequence encoding
c 33	28.6	7.6	4690	1	I94812	Rat penile neurona
34	28.4	7.5	34720	1	V30458_5	Continuation (6 of
35	28.4	7.5	3241	1	V72660	Caenorhabditis ele
36	28.4	7.5	638	1	X21014	Polynucleotide seq
37	28.4	7.5	65165	1	V30459_5	Continuation (6 of
c 38	28.2	7.5	4171	1	Q74082	Oncostatin M recep
c 39	28.2	7.5	4626	1	V20444	Human c-met oncoge
40	28.2	7.5	2400	1	V94239	Human SIAH-130 enc
c 41	28.2	7.5	941	1	V94414	M. musculus predic
c 42	28.2	7.5	941	1	V94419	Human secreted pro
43	28.2	7.5	1038	1	X [illegible]	Human secreted pro
c 44	28	7.4	1287	1	I28947	Synthetic N-C. New
c 45	28	7.4	1290	1	I78692	Sequence encoding

ALIGNMENTS

RESULT 1
X39873
ID X39873 standard; DNA; 1227 BP.
AC X39873;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated antig.
KW Cancer associated antigen; diagnosis; [illegible] human;
KW breast cancer; colon cancer; gastric cancer; [illegible] cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN WO9904200-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14879.
PR 22-JUN-1998; US-102922.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, Chen M, Ikeda Y [illegible]
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132148/11.
PT New isolated cancer associated nucle[illegible] acids and polypeptides -
PT isolated using sera from cancer patients [illegible] breast, prostate
PT for the diagnosis, monitoring or treatment of tumours
PS Claim 67; Page 579-580; 787pp; English.
CC The invention relates to a method for diagnosing [illegible] characterised
CC by expression of a human ca[illegible] associated [illegible] for by
CC a nucleic acid molecule (NAM). The method [illegible] using a
CC biological sample isolated [illegible] a NAM which specifically
CC binds to the NAM, an expres[illegible] product [illegible] expression
CC product complexed with an HLA molecule and [illegible] The
CC interaction between the ab[illegible] and the NAM or [illegible] is [illegible] as a
CC determination of the disord[illegible] The p[illegible] are [illegible] used in
CC the diagnosis, monitoring, research, or treatment of [illegible]
CC characterised by the expres[illegible] of [illegible] tumour associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigens predicted [illegible] in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 1227 BP; 388 A; 270 C; 344 G; 225 T;

Query Match 70.5%; Score 266.4; DB 1; Length 1227;
Best Local Similarity 99.3%; Pred. No. 6e-71;
Matches 278; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 98 gattactgttcagcatgtacgcaattgcgtccttcgcgtttgcgtatgtt 147
Db 880 GACTACTGTTCAGCATGTACGCAATTGCGTCCTTCGCGTTTGCGTATGTT 929

Qy 148 gtgctgcagacatgaggcctgatggcctgattcttgtttgtgtatgtgtttat 197
Db 930 GTGCTGCAGACATGAGGCCTGAT[illegible] 979

Qy 198 ggagaatgaaggctcactggcctgattatgttgtttgtgtatgaaatttatat 247
Db 1030 GGAGAATGAAGGTCACTGGCCCTGAT[illegible] 1079

Qy 248 ttgtgttgatgatgaagttgg[illegible] 306

```
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    1060 TTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTT 1119

Qy     307 gataccccccgatgaagcagattcttttgagtactgggcc 346
          |||||||||||||||||||||||||||||||||||||||||
Db    1120 GATACCCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCC 1159


RESULT   2
X39876
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis, research, treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102222.
PR   17-JUL-1997; US-896194.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determinant of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of nucleic acid associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    209 G;    184 T;


  Query Match             70.5%;  Score 266.4;  DB 1;  Length 816;
  Best Local Similarity   99.3%;  Pred. No. 6.6e-79;
  Matches  278;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      68 gactactgtcaggaggatgtaggaatggacat-gcttttgaaagagggcgtgctgagtccca 127
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     469 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCA 528

Qy     128 gtgctccagacatgaaggcctgaacctcctaattctctggatcttaatgacactcatcctc 187
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     529 GTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTC 588

Qy     188 ggaggaatcaagctccacagcccccaaatatcaat-tttcttgtgaccaaagtgaaggatcta 246
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     589 GGAGGAATCAAGCTCCACAGCCCCCAAATATCAATCTTTCTCTGTGACCAAAGTGAAGGATCTA 648

Qy     248 ttctctctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaacttc 306
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     649 TTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTC 708


Qy     307 gataccccccgatgaagcagattcttttgagtactgggcc 346
          |||||||||||||||||||||||||||||||||||||||||
Db     709 GATACCCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCC 748


RESULT   3
X29771
ID   X29771 standard; DNA; 1700 BP.
AC   X29771;
DT   22-JUN-1999  (first entry)
DE   L.lactis HsdM gene #2.
KW   HsdR; HsdM; HsdS; restriction/modification; bacteriophage; resistance;
KW   lactic acid bacterium; ds.
OS   Lactococcus lactis.
PN   FR2767831-A1.
PD   05-MAR-1999.
PF   02-SEP-1997; 010885.
PR   02-SEP-1997; FR-010885
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PI   Chopin MC, Clier F, Ehrlich SD, Gautier M, Schouler C;
DR   WPI; 99-183265/16
DR   P-PSDB; Y04661.
PT   Bacteriophage resistance mechanism subunit polypeptides - of lactic
PT   acid bacteria, especially Lactococcus lactis
PS   Claim 5; Page 39-41; 65pp; French.
CC   The invention relates to polypeptides that constitute the HsdR, HsdM or
CC   HsdS subunit of a R/M [restriction/modification] type bacteriophage
CC   resistance mechanism active against the phages of lactic acid bacteria,
CC   especially Lactococcus lactis. This sequence represents the gene
CC   encoding the Lactococcus lactis HsdM subunit. Nucleic acids encoding
CC   the HsdR, HsdM or HsdS fragment can be used for expression of at least
CC   one bacteriophage resistance mechanism in a lactic acid bacterium.
SQ   Sequence   1700 BP;    541 A;    299 C;    368 G;    492 T;


  Query Match             9.0%;  Score 34.2;  DB 1;  Length 1700;
  Best Local Similarity   54.3%;  Pred. No. 0.11;
  Matches   69;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     212 atatcaatctttctctggaccaaagtaaaagatctattttctctgtgatgatgaacttc 271
          ||| | | ||||| | || |||  ||| | ||||  |  |  |  |  |  |  |  |
Db    1360 ATAAAATTGTTTCTATCCACAAAAAAGCAGAAGATATTGAAAGATACTTTATTTTTAA 1419

Qy     272 gtccagatgaaattgacatcaatgtggatgaacttcgatagcccccatgaagcagattcttt 331
          ||  ||||| || |  |||  |  | | | | |  |  |  |  |  |  |  |  |  |
Db    1420 GTTTTGATGAGATCCAAGAAAATGACTTTAACTTAAATATCCCTGTTAGTTATATTT 1479

Qy     332 ttgagta 338
          ||||| |
Db    1480 TTGAGGA 1486


RESULT   4
T38228
ID   T38228 standard; cDNA; 1386 BP.
AC   T38228;
DT   18-DEC-1996  (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA.
KW   Apoptosis-regulating protein; Nip2; adenovirus E1B 19K protein;
KW   Bcl-2; oncogene; cell death; cancer; tumour; immune disorder;
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214   1158
FT                   /*tag= a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV SI LOUIS.
PI   Chinnadurai G;
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998.
```

```
PT   Nucleic acids encoding apoptosis regulating proteins - useful for
PT   diagnosing and treating immune disorders, malignancies, etc.
PS   Claim 1; Page 39-41; 60pp; English.
CC   6 cDNA clones (T29227-32) respectively code for novel apoptosis-
CC   regulating proteins Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13
CC   (W00997-W01002).  Nip clones were isolated from human cDNA
CC   libraries using a yeast two-hybrid stategy with adenovirus E1B
CC   19K protein (W01010) as bait; Bip clones were isolated using
CC   Bcl-2 oncoprotein (W01018) as bait.  E1B 19K and Bcl-2 both protect
CC   cells against apoptosis.  Nip1, Nip2, Nip3 bind to 19K, Bip1A and
CC   Bip5 bind to 19K and Bcl-2, and Bip5 binds to Bcl-2 alone.  Nucleic
CC   acids can be used in diagnostic assays, to produce recombinant
CC   apoptosis-regulating proteins, or to regulate cell death by
CC   increasing or suppressing expression of an apoptosis-blocking gene
CC   such as bcl-2.
SQ   Sequence   1386 BP;    428 A;    249 C;    336 G;    373 T;

  Query Match              9.0%;  Score 34;  DB 1;  Length 1386;
  Best Local Similarity   61.1%;  Pred. No. 0.11;
  Matches   55;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     189 gagaatcaagctcacagcccaaatatcaatctttctctggaccaaagtgaaggatctat 248
           ||||| || || |  || ||| | || |   |  | ||||| |  ||||| || ||| |
Db     360 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGCTCTGT 419

Qy     249 tctctctgatgataacttggacagtccaaga 278
           | || ||||||    || ||| |  |
Db     420 ATTGTCAGATGATTTGGATGAAAGTGGGGA 449


RESULT   5
X04120
ID   X04120 standard; cDNA; 1386 BP.
AC   X04120;
DT   13-APR-1999  (first entry)
DE   Human Nip2 encoding cDNA.
KW   Human, Nip1, Nip2, Nip3, Bip1A, Bip5, Bip13; adenovirus; cell death;
KW   viral infection; Bcl-2, protooncogene, mutational analysis, apoptosis,
KW   E1B 19K protein; cell survival regulation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             214   1158
FT                   /*tag=  a
PN   US5858678-A.
PD   12-JAN-1999.
PF   21-MAR-1995; 408095.
PR   02-AUG-1994; US-284139.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 99-152099/13.
DR   P-PSDB; W94324.
PT   Polypeptides that bind to anti-apoptotic proteins - useful for
PT   protecting against cell death induced by viral infection and to
PT   modulate response to physical and chemical stimuli
PS   Claim 2; Column 51-54; 41pp; English.
CC   The present invention describes: (1) a method for regulating cell death,
CC   comprising exposing an isolated cell to a polypeptide selected from
CC   Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13; (2) a method for neutralising
CC   the activity of the adenovirus E1B 19 kD protein, the Bcl-2 protein or
CC   the BHRF-1 protein, comprising exposing an isolated cell to a
CC   polypeptide as in (1); and (3) a method for detecting molecules that
CC   bind to at least one polypeptide as in (1), comprising lysing cells,
CC   exposing the lysate to the polypeptide and detecting any molecule-
CC   polypeptide aggregates. The methods are useful for providing proteins
CC   able to bind to other proteins known to regulate cell survival e.g. it
CC   is known that E1B 19K protein provides a survival function similar to
CC   the cellular protooncogene bcl-2 gene product which is able to block
CC   apoptosis in haematopoietic B and T cells. The present sequence encodes
CC   human Nip2, given in the present invention.
SQ   Sequence   1386 BP;    428 A;    249 C;    336 G,    373 T,

  Query Match              9.0%;  Score 34;  DB 1;  Length 1386;
  Best Local Similarity   61.1%;  Pred. No. 0.11;
  Matches   55;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     189 gagaatcaagctcacagcccaaatatcaatctttctctggaccaaagtgaaggatctat 248
           ||||| || || |  || ||| | || |   |  | ||||| |  ||||| || ||| |
Db     360 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGCTCTGT 419

Qy     249 tctctctgatgataacttggacagtccaaga 278
           | || ||||||    || ||| |  |
Db     420 ATTGTCAGATGATTTGGATGAAAGTGGGGA 449


RESULT   6
I61583
ID   I61583 standard; DNA; 3001 BP.
AC   I61583;
DT   29-OCT-1997  (first entry)
DE   Malolactic enzyme and malate permease genes of Leuconostoc oenos.
KW   Fermentation; wine making; wine must; conversion; lactic acid; MLEA;
KW   MLEP; malic acid; ds.
OS   Leuconostoc oenos strain Lo 84.13.
FH   Key             Location/Qualifiers
FT   cds             364. .1989
FT                   /*tag=  a
FT                   /product= malolactic_enzyme
FT   cds             1989. .2933
FT                   /*tag=  b
FT                   /product= malate_permease
FT                   /note= "The first nucleotide of this coding region
FT                   overlaps with the last nucleotide of the
FT                   termination codon for the malolactic enzyme
FT                   gene"
PN   FR2736652-A1.
PD   17-JAN-1997.
PF   13-JUL-1995; 008705.
PR   13-JUL-1995; FR-008705.
PA   (UYBO-) UNIV BOURGOGNE.
PI   Cavin JF,  Divies C,  Guzzo J,  Labarre C;
DR   WPI; 97-122440/12.
DR   P-PSDB; W13542, W13543.
PT   DNA encoding malolactic enzyme and/or malate permease enzyme - is
PT   Leuconostoc oenos, useful in wine production
PS   Claim 1; Fig 2; 33pp; French.
CC   Two degenerate oligonucleotides having the sequences given in
CC   I61584 and I61585 were synthesised based on highly conserved regions
CC   in the sequence of enzymes which use NADH as co-factor and use malic
CC   acid as substrate. The primer pair was used in a PCR amplification
CC   to obtain a 324 bp fragment of the malolactic enzyme gene from
CC   Leuconostoc oenos strain Lo 84.13. Using the amplified fragment as
CC   a probe, seven positive clones were identified in a plasmid cDNA
CC   library. Further analysis of two of the clones revealed that they
CC   contained a region of 3001 bp which contained two open reading
CC   frames (ORFs) (i.e. the present sequence). The first ORF encodes a
CC   malolactic enzyme and the second ORF, which overlaps the last residue
CC   of the first ORF, encodes malate permease. Microorganisms (especially
CC   Brewer's yeast, Saccharomyces cerevisiae) transformed with these genes
CC   can be used to convert malic acid present in wine must into lactic acid.
CC   By providing malate permease, the  malolactic activity of the
CC   microorganism is not limited by the speed of penetration of L-malate
CC   into the yeast cells.
SQ   Sequence   3001 BP;    814 A;    528 C;    647 G;    1012 T;

  Query Match              8.2%;  Score 31;  DB 1;  Length 3001;
  Best Local Similarity   53.8%;  Pred. No. 1.6;
  Matches   64;  Conservative    7;  Mismatches   [illegible];  Indels    [illegible];  Gaps    0;

Qy     217 aatcttttctctggaccaaagtgaaggatctattctctctgatgataacttggacagtcc 276
           |  ||||| |    | |||  |
Db     592 AAGCTTTTTTCTCAACATGTTGTTGAATTA-ACCAATTGCTAATAATTGATTAGTTTCC 651
```

```
Qy     277 gatgaaatttgaatctaatgtggatgaaacttgatacccccgatgaaagcagattcttttga 335
           ||| ||||||| | | ||| | | |  |  || | | || | || ||| ||
Db     652 GATACAATTGAAAAATTATTCGGAATTATTTGTGAACCGCAAGGTGCCGCTTTTTTGGA 710


RESULT   7
V31262/c
ID   V31262 standard; DNA; 824 BP.
AC   V31262;
DT   01-OCT-1998 (first entry)
DE   E. coli J96 pathogenicity island contig #70.
KW   PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
KW   PAI V; pheV; vaccine; protective immune response; ds.
OS   Escherichia coli.
PN   WO9822575-A2.
PD   28-MAY-1998.
PF   21-NOV-1997; U21347.
PR   14-OCT-1997; US-061958.
PR   22-NOV-1996; US-031626.
PA   (HUMA-) HUMAN GENOME SCI INC
PA   (UYWI-) UNIV WISCONSIN.
PI   Choi GH, Dillon PJ, Welch RA.
DR   WPI; 98-312461/27.
PT   New isolated uropathogenic E. coli nucleotide sequences - used to
PT   develop products for the detection of pathogenic E. coli and to
PT   elicit an immune response to pathogenic E. coli
PS   Claim 21; Page 190; 250pp; English.
CC   This sequence represents a E. coli strain J96 contig containing
CC   pathogenicity island (PAI) sequences, and represents a nucleic acid
CC   molecule of the invention. PAIs are large fragments of DNA which comprise
CC   pathogenicity determinants. The sequences of the invention are taken from
CC   PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
CC   on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC   approximately 94 min (at pheR) on the E. coli chromosome and is
CC   approximately 160 kb in size. Antibodies specific to the proteins encoded
CC   by the PAI open reading frames of the invention can be used in kits to
CC   detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC   a protective immune response in an animal to the uropathogenic E. coli
CC   strain J96.
SQ   Sequence   824 BP;    212 A;    222 C;    131 G;    257 T;


  Query Match             8.2%;  Score 31;  DB 1;  Length 824;
  Best Local Similarity   47.2%;  Pred. No. 0.89;
  Matches   94;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     129 tgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcctcg 188
           | | |  |  | ||  |||||    | |  || | |  | |   |||||||
Db     791 TTCAGTTTGAAAAAGAACTTGAAAAAACTCAATCTGGAGTTATGAATTCACTCATATCAA 732

Qy     189 gagaatcaaatcaacagcccaaatatcgat tttctctggacaaagtgaaggatctat 248
           | |  ||  | ||||   |  ||| |    |||  ||||  ||  |   ||  | |
Db     731 TCGTAAAAATTTAACAGAGACTCATAACTTATTTCCGTGGAAGAACTTATTGGGACCAC 672

Qy     249 tctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttga 308
           |  | ||| |   || | ||||  | ||  |  | || |||  |  ||   |||
Db     671 GCGTGCCGACAAGTTTTTTGGCAGTAAAATTTCTGATATCTTCCATGGCGCGGATGGTGA 612

Qy     309 tacccccgatgaaagcagat 327
           |  ||    |||| | ||
Db     611 TGACCATATAGAAGGAAAT 593


RESULT   8
V90545/c
ID   V90545 standard; DNA; 1610 BP.
AC   V90545;
DT   18-FEB-1999 (first entry)
DE   Nucleotide sequence from clone G3a.
KW   Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW   peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS   Helicobacter pylori.
PN   WO9849314-A2.
PD   05-NOV-1998.
PF   27-APR-1998; U08487.
PP   14-OCT-1997; US-061958.
PR   25-APR-1997; US-045107.
PA   (GENE-) GENELABS TECHNOLOGIES INC.
PI   Chow TP, Fry KE, Lim MY, McAtee CP;
DR   WPI; 99-009433/01.
PT   New Helicobacter pylori antigens and related nucleic acid sequences
PT   - useful in serological diagnosis and protective vaccines, providing
PT   long-lasting immune response
PS   Claim 27; Page 95; 402pp; English.
CC   The present sequence encodes a Helicobacter pylori antigenic protein
CC   that is characterised by immunoreactivity with H. pylori-positive
CC   antisera. The proteins are highly immunogenic and induce a long-lasting
CC   immune response that persists even after antimicrobial treatment. In
CC   antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC   highly sensitive and specific. The specification also describes 64
CC   previously unrecognised immunogenic cluster families. H. pylori antigens
CC   are used to detect H. pylori-specific antibodies for diagnosis of
CC   infection or to confirm eradication of infection, and in vaccines to
CC   protect against H. pylori infection and related diseases (gastritis,
CC   peptic ulcer, gastric adenocarcinoma/lymphoma).
SQ   Sequence   1610 BP;    545 A;    296 C;    328 G;    441 T;


  Query Match             8.0%;  Score 30.4;  DB 1;  Length 1610;
  Best Local Similarity   59.1%;  Pred. No. 1.9;
  Matches   52;  Conservative    0;  Mismatches   36;  Indels    1;  Gaps    0;

Qy     246 tattctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaact 305
           | || ||| | ||| |  | | |  ||   | |||||   |||||| | | |||
Db     722 TTTTTTCTTTCATGTTGGCGTTGGTAGCGGAAATGAAACCAACATCAATTTTAACGTTT 663

Qy     306 tgataccccccgatgaaagcagattctttt 333
           |  || |||| | |     ||| |||
Db     662 TATTATCCCCAAGGCGCGTGATTTCTTT 635


RESULT   9
X13324
ID   X13324 standard; DNA; 3658 BP.
AC   X13324;
DT   19-MAR-1999 (first entry)
DE   Enterococcus faecalis genome contig; SEQ ID NO:387.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection; in;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; U08985.
PF   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PP   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 1569-1571; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
```

```
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence   3658 BP;    1069 A;    753 C;    670 G;    1164 T;


  Query Match             8.0%;  Score 30.2;  DB 1;  Length 3658;
  Best Local Similarity   56.6%;  Pred. No. 3.3;
  Matches   56;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     255 tgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgataccc 314
           |||||  ||  | | ||||  |   | |||  |   | |  ||    ||| |||||
Db     515 TTATGAAAAACTCGTCAGTTCCATTCAAAAGGATGTGACGGTTTTAGAAATTGATTTATA 574

Qy     315 cgatgaagcagattcttttgagtacactggccatggtaa 353
           | || || ||| || || || ||    | |  |||| ||
Db     575 CAATCAAACAGGTTGTTATGGGTTGTATCGAAATGGCAA 613


RESULT  10
V72213/c
ID   V72213 standard; DNA; 3190 BP.
AC   V72213;
DT   27-APR-1999  (first entry)
DE   Human NT2-1A DNA.
KW   NBC; sodium bicarbonate transporter family; immunoassay; pH regulation;
KW   treatment; water retention; blood pressure; acidosis; inflammation;
KW   cell proliferation; cancer; sperm activation; inactivation; epilepsy;
KW   hydroencephaly; glaucoma; colitis; human; NT2-1A; ss.
OS   Homo sapiens.
PN   WO9853067-A1.
PD   26-NOV-1998.
PF   20-MAY-1998; 98WO-US010297.
PR   20-MAY-1997; 97US-047131.
PA   (UYYA ) UNIV YALE.
PI   Aalkjaer C, Bevensee MO, Biemesderfer D, Boron WF,
PI   Choi I, Davis BA, Grichtchenko II, Romero MF, Schmitt BM,
PI   Sussman CR;
DR   WPI; 99-059743/05.
PT   New nucleic acid molecules encoding proteins of the Sodium
PT   Bicarbonate Cotransporter (NBC) family - useful for identifying
PT   agents that agonise or antagonise NBC activity and treating
PT   disorders mediated by NBC
PS   Example 5; Fig 18; 138pp; English.
CC   This sequence encodes the human NT2-1A protein which is used in a method
CC   resulting in the isolation of a novel sodium bicarbonate transporter.
CC   NBC proteins and nucleic acid sequences may be used to treat pathological
CC   processes including water retention, increased blood pressure, chronic
CC   respiratory and metabolic acidosis, inflammation, cell proliferation,
CC   cancer, sperm activation or inactivation, hydroencephaly, epilepsy,
CC   glaucoma and colitis. Members of the NBC family of proteins can be used
CC   (i) as a target to identify agents that block or stimulate NBC mediated
CC   pH regulation, (ii) as a target or bait to identify and isolate binding
CC   partners that bind NBC, (iii) in methods to identify agents that block or
CC   stimulate activity of an NBC protein, and (iv) as a target to assay for
CC   NBC-mediated activity. Anti-NBC antibodies are also useful as modulators
CC   of NBC activity, useful in immunoassays for detecting NBC expression or
CC   activity and for purifying an NBC protein.
SQ   Sequence   3190 BP;    921 A;    585 C;    706 G;    977 T;


  Query Match             8.0%;  Score 30.2;  DB 1;  Length 3190;
  Best Local Similarity   60.2%;  Pred. No. 3.1;
  Matches   50;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     255 tgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgataccc 314
           |||||| |     | | |||  ||||  |   ||    |  |||||||||||   |  |
Db      94 TGATGATGAAGAACATATTCCCCATGATTCTTCACGGAAATGGATGAACTGTGTTACAG 25

Qy     315 cgatgaagcagattcttttgagt 337
           | ||  |||||  ||||||  |||
Db      24 CCATCTATCATTTCTTTCTTAGT 2


RESULT  11
V72214/c
ID   V72214 standard; DNA; 4417 BP.
AC   V72214;
DT   27-APR-1999  (first entry)
DE   Human NT2-2A DNA.
KW   NBC; sodium bicarbonate transporter family; immunoassay; pH regulation;
KW   treatment; water retention; blood pressure; acidosis; inflammation;
KW   cell proliferation; cancer; sperm activation; inactivation; epilepsy;
KW   hydroencephaly; glaucoma; colitis; human; NT2-2A; ss.
OS   Homo sapiens.
PN   WO9853067-A1.
PD   26-NOV-1998.
PF   20-MAY-1998; 98WO-US010297.
PR   20-MAY-1997; 97US-047131.
PA   (UYYA ) UNIV YALE.
PI   Aalkjaer C, Bevensee MO, Biemesderfer D, Boron WF,
PI   Choi I, Davis BA, Grichtchenko II, Romero MF, Schmitt BM,
PI   Sussman CR;
DR   WPI; 99-059743/05.
DR   P-PSDB; W90274.
PT   New nucleic acid molecules encoding proteins of the Sodium
PT   Bicarbonate Cotransporter (NBC) family - useful for identifying
PT   agents that agonise or antagonise NBC activity and treating
PT   disorders mediated by NBC
PS   Example 5; Fig 19A; 138pp; English.
CC   This sequence encodes the human NT2-2A protein which is used in a method
CC   resulting in the isolation of a novel sodium bicarbonate transporter.
CC   NBC proteins and nucleic acid sequences may be used to treat pathological
CC   processes including water retention, increased blood pressure, chronic
CC   respiratory and metabolic acidosis, inflammation, cell proliferation,
CC   cancer, sperm activation or inactivation, hydroencephaly, epilepsy,
CC   glaucoma and colitis. Members of the NBC family of proteins can be used
CC   (i) as a target to identify agents that block or stimulate NBC mediated
CC   pH regulation, (ii) as a target or bait to identify and isolate binding
CC   partners that bind NBC, (iii) in methods to identify agents that block or
CC   stimulate activity of an NBC protein, and (iv) as a target to assay for
CC   NBC-mediated activity. Anti-NBC antibodies are also useful as modulators
CC   of NBC activity, useful in immunoassays for detecting NBC expression or
CC   activity and for purifying an NBC protein.
SQ   Sequence   4417 BP;    1298 A;    [illegible] C;    [illegible] G;    [illegible] T;


  Query Match             8.0%;  Score 30.2;  DB 1;  Length 4417;
  Best Local Similarity   60.2%;  Pred. No. 3.6;
  Matches   50;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy     255 tgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgataccc 314
           |||||| |     | | |||  ||||  |   ||    |  |||||||||||   |  |
Db      83 TGATGATGAAGAACATATTCCCCATGATTCTTCACGGAAATGGATGAACTGTGTTACAG 24

Qy     315 cgatgaagcagattcttttgagt 337
           | ||  |||||  ||||||  |||
Db      23 CCATCTATCATTTCTTTCTTAGT 1


RESULT  12
Q10263
ID   Q10263 standard; DNA; 4383 BP.
AC   Q10263;
DT   04-APR-1991  (first entry)
DE   pZ130 contg. Calgene lambda 111 genomic clone.
KW   Ovary tissue transcriptional factor; DNA construct; promo;
KW   clone pZ130; ss.
OS   Lycopersicon esculentum UC82B.
FH   Key             Location/Qualifiers
FT   intron          2702..2921
FT                   /*tag= a
FT   misc_rna        2599..3382
FT                   /*tag= b
FT                   /label= pZ130
FT                   /note= "overlap"
FT   misc_feature    2567..2567
```

```
FT                   /*tag=  c
FT                   /label= transcription_start
PN   EP-409629-A.
PD   23-JAN-1991.
PF   19-JUL-1990; 307926.
PR   19-JUL-1989; US-382519.
PA   (CALG-) CALGENE INC.
PI   Martineau B, Houck CM,
DR   WPI; 91-024191/04.
FT   New ovary tissue transcriptional factors - modify transcription
PT   in tomato plant ovaries for use as mol. probes
PS   Disclosure; Fig 2; 21pp; English.
CC   To prepare pCGN2901, Calgene Lambda 140 was digested with SalI
CC   and the resulting fragment which contains the pZ7-hybridising
CC   region was inserted into pCGN2015, at the unique SalI site.
CC   pCGN2902 contains the other SalI fragment (non-pZ7-hybridising)
CC   of the pZ130 genome derived from SalI digestion of
CC   Calgene Lambda 140 also put into a pCGN2015 construct.
CC   Plasmid DNA isolated from pCGN2901 was used in the construction
CC   of a pZ130 Promoter Cassette.  The promoter is ovary-specific
CC   and wound-inducible in leaf cells.
CC   See also Q10262-64.
SQ   Sequence   4383 BP;    1539 A;    620 C;    610 G;    1606 T;

  Query Match             7.9%;  Score 30;  DB 1,  Length 4383,
  Best Local Similarity   55.9%;  Pred. No. 4.2;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     152 ctcctaattctctggatcttaatgacacctcatcctcggagaatcaagctcacagccccaa 211
           ||| || ||| ||   | | ||| || | |||  |  || | |     ||||| | | ||
Db    2517 CTCTTATTTCACTCCCTATAAATAACTCCCATTGTCTGATATTTTTATTCACAACTCTAA 2576

Qy     212 atatcaatctttctctggaccaaagtgaaggatctattctct 253
           |  ||||||||||    |  |||   ||  |     | |||
Db    2577 CTTACAATCTTTCTTATTATTAAAAAAAACAAAAACATTTCT 2619


RESULT  13
Q10319
ID   Q10319 standard; cDNA; 4383 BP.
AC   Q10319;
DT   04-APR-1991  (first entry)
DE   Calgene lambda 140 genomic clone.
KW   pZ130; cytokinin; ds
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_rna        2567..3382
FT                   /*tag=  a
FT                   /note= "Sequence homologous to pZ130 transcript"
FT   misc_rna        2599..3382
FT                   /*tag=  b
FT                   /note= "Sequence homologous to pZ130 CDS"
FT   misc_rna        2702..2921
FT                   /*tag=  c
FT                   /note= "Sequence homologous to pZ130 intron"
PN   EP-409628-A.
PD   23-JAN-1991.
PF   19-JUL-1990; 307925.
PR   19-JUL-1989; US-382802.
PA   (CALG-) CALGENE INC.
PI   Houck CM, Pear JR, Martineau B, Hiatt W;
DR   WPI; 91-024190/04.
PT   Modulating endogenous cytokinin levels - regulatory regions are
PT   transformed into plant cells e.g. fruit to modify phenotype
PS   Disclosure; Fig 3; 39pp; English.
CC   The sequence encodes an enzyme in the cytokinin biosynthetic pathway,
CC   derived from Z130, and useful in modulating a transformed plant's
CC   phenotype eg. fruit maturation, ripening etc.
SQ   Sequence   4383 BP;    1539 A;    620 C;    610 G;    1606 T;

  Query Match             7.9%;  Score 30;  DB 1;  Length 4383;
  Best Local Similarity   55.9%;  Pred. No. 4.2;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     152 ctcctaattctctggatcttaatgacacctcatcctcggagaatcaagctcacagccccaa 211
           ||| || ||| ||   | | ||| || | |||  |  || | |     ||||| | | ||
Db    2517 CTCTTATTTCACTCCCTATAAATAACTCCCATTGTCTGATATTTTTATTCACAACTCTAA 2576

Qy     212 atatcaatctttctctggaccaaagtgaaggatctattctct 253
           |  ||||||||||    |  |||   ||  |     | |||
Db    2577 CTTACAATCTTTCTTATTATTAAAAAAAACAAAAACATTTCT 2619


RESULT  14
Q35143
ID   Q35143 standard; cDNA; 4383 BP.
AC   Q35143;
DT   24-MAY-1993  (first entry)
DE   Calgene lambda 140/pZ130 DNA including pZ130 gene.
KW   Calgene lambda 140; pZ130; expression; cytokinin; plant; ss.
FH   Key             Location/Qualifiers
FT   intron          2702..2921
FT                   /*tag=  a
FT   misc_feature    2599..3382
FT                   /*tag=  b
FT                   /note= "overlap between lambda 140 genomic clone
FT                   with pZ130 cDNA clone"
FT   misc_rna        2567
FT                   /*tag=  c
FT                   /note= "start of pZ130 gene transcript"
PN   US5177307-A.
PD   05-JAN-1993.
PF   26-MAY-1987; 054369.
PR   26-MAY-1987; US-054369.
PR   15-MAR-1988; US-168190.
PR   29-APR-1988; US-188361.
PR   19-JUL-1989; US-382802
PR   17-JUL-1990; US-554196.
PA   (CALJ ) CALGENE INC.
PI   Hiatt W,  Houck CM,  Martineau BM,  Pear JR;
DR   WPI; 93-035826/04.
PT   New plant expression cassettes - contg. DNA encoding an enzyme in
PT   a cytokinin metabolic pathway to modify cytokinin levels in
PT   plants
PS   Disclosure; Fig 3; 29pp; English.
CC   The sequence shows the complete sequence of the region of Calgene
CC   lambda 140 genomic clone that overlaps with the pZ130 cDNA clone and
CC   a partial sequence of regions 5' and 3' to that region.  It is
CC   possible that small gaps in the sequence may exist upstream from the
CC   SalI site bound at position 808.  The clone is used in the
CC   construction of a novel expression cassette comprising in the 5' 3'
CC   direction of transcription, as operably joined components: a
CC   transcriptional and translational initiation region from Z130, a DNA
CC   sequence encoding an enzyme in a cytokinin metabolic pathway and a
CC   transcription termination region.  The enzyme is pref. DMA
CC   transferase.  The expression cassette can be used to modify a plant
CC   phenotype by modifying cytokinin levels during a specific time and/or
CC   in specific tissues during the development of a plant.  Control of
CC   cytokinin concns. can be used to increase the total solids content of
CC   ripe fruit, improve retention of early fruit and alter the time of
CC   onset of ripening.        See also Q25144-5.
SQ   Sequence   4383 BP;    1540 A;    628 C;    606 G;    1609 T;

  Query Match             7.9%;  Score 30;  DB 1;  Length 4383;
  Best Local Similarity   55.9%;  Pred. No. 4.2;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     152 ctcctaattctctggatcttaatgacacctcatcctcggagaatcaagctcacagccccaa 211
           ||| || ||| ||   | | ||| || | |||  |  || | |     ||||| | | ||
Db    2517 CTCTTATTTCACTCCCTATAAATAACTCCCATTGTCTGATATTTTTATTCACAACTCTAA 2576

Qy     212 atatcaatctttctctggaccaaagtgaaggatctattctct 253
           |  ||||||| ||    |  |||   ||  |     | |||
```

```
Db    2577 CTTACAATCTTTCTTATTATTAAAAAAAACAAAAACATTTCT 2618


RESULT  15
Q34941
ID   Q34941 standard; DNA; 4383 BP.
AC   Q34941;
DT   19-MAY-1993  (first entry)
DE   Calgene Lambda 140 genomic clone.
KW   cDNA; clone pZ130; anthesis, tomato, ovary, integumen, outer pericarp,
KW   fruit; developememt; transcription; initiation; region; modulation;
KW   ovary-specific; endogenous; fruit product; exogenous; phenotype;
KW   lycopersicon esculentum; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_rna        2599. .3382
FT                   /*tag=  a
FT                   /note= "Interrupted pZ130 cDNA clone"
FT   intron          2702. .2921
FT                   /*tag=  b
FT   misc_signal     2567
FT                   /*tag=  c
FT                   /function= Start of pZ130 gene transcript
PN   US5175095-A.
PD   29-DEC-1992.
PF   19-JUL-1989; 382518.
PR   19-JUL-1989; US-382518.
PR   17-JUL-1990; US-554195.
PA   (CALJ ) CALGENE INC.
PI   Houck CM, Martineau BM;
DR   WPI; 93-026940/03.
PT   DNA constructs contg. tomato pZ130 transcriptional initiation
PT   region - useful for modulation of endogenous fruit prods. and for
PT   prodn. of exogenous prods.
PS   Disclosure; Fig 2; 18pp; English.
CC   The sequence given shows the sequence of the region of the Calgene
CC   Lambda 140 genomic clone which overlaps with the pZ130 cDNA clone and
CC   a partial sequence 5' and 3 to that region.  The pZ130 sequence is
CC   expressed during the early stages of anthesis in tomatos.  The message
CC   is expressed in ovary integumen and ovary outer pericarp tissue.  It
CC   is not readily detectable in other tissues or at other stages of fruit
CC   developement.  The transcription initiation region associated with
CC   this gene is therefore considered to be ovary-specific.  The actual
CC   function of the pZ130 polypeptide is unknown.  The transcription
CC   initiation region can be used for modulation of endogenous fruit
CC   products, for production of exogenous products and for modification
CC   of the phenotype of fruit and fruit products.
SQ   Sequence   4383 BP;    1539 A;    920 C;    810 G;    1606 T,


  Query Match             7.9%;  Score 30;  DB 1;  Length 4383;
  Best Local Similarity   55.9%;  Pred. No. 4.2;
  Matches   57;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     152 ctcctaattctctggatcttaatgacactcatcctcggagaatcaagctcacagccccaa 211
           ||| || ||| ||   | | ||| || ' '|'  |  || | |     |||||  | | ||
Db    2517 CTCTTATTTCACTCCCTATAAATAACTCCCATTGTGTGATATTTTTATTCACAACTCTAA 2576

Qy     212 atatcaatctttctctggaccaaagtgaaggatctattctct 253
           |  '|||||||||    |  |||   ||       | |||
Db    2577 CTTACAATCTTTCTTATTATTAAAAAAAACAAAAACATTTCT 2618


Search completed: February  8, 2000, 20:54:10
Job time: 1498 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 07:52:19 ; Search time 5888.89 Seconds
(without alignments)
-498.081 Million cell updates/sec

Title: US-09-174-937-6
Perfect score: 966
Sequence: 1 atggaggaggagacagagtt.....tgaagctgaaagaaaagcct 966

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, 1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold : 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	901.8	93.4	5654	9	AB002365	AB002365 Human mRN
2	294.9	29.5	1386	5	AR028453	AR028453 Sequence
3	284.8	29.5	2382	11	HS215171	H215171 Homo sapien
4	283	29.3	2213	12	AF115207	AF115207 Mus muscu
5	202.4	21.0	169738	44	AC003656	AC003656 Homo sapi
c 6	172.8	17.9	169798	44	AC003656	AC003656 Homo sapi
c 7	92	9.5	179301	41	AC011488	AC011488 Homo sapi
c 8	85.6	8.9	78170	43	AC005096	AC005096 Drosophil
9	60	6.2	1430	9	HSRHCGAPX	72924 H.sapiens r
10	60	6.2	3345	9	HSCDC42	[illegible] human CDC42
c 11	56	5.8	58386	42	AC012249	AC012249 Homo sapi
12	54	5.6	125121	42	AC011451	AC011451 Drosophil
13	44.6	4.6	1378	10	HSCDC794	M61794 Human CDC42
c 14	42.6	4.4	7218	5	T66484	T66484 Sequence 14
c 15	41.4	4.3	86004	11	HS51023	AC003512 Human DNA
c 16	38.8	4.0	101933	8	AC004146	AC004146 Arabidops
17	38.6	4.0	200058	42	AC016431	AC016431 Homo sapi
18	38.4	4.0	12999	2	AE000761	AE000761 Aquifex a
19	38.2	4.0	161100	44	AC013657	AC013657 Homo sapi
20	38	3.9	108056	8	ATAC000577	AC000577 Arabidops
21	37.8	3.9	1251	10	HSU77616	U77616 Human achae
c 22	37.2	3.9	28771	10	AB014086	AB014086 Homo sapi
c 23	37.2	3.9	100000	10	AP000516	AP000516 Homo sapi
c 24	37.2	3.9	40057	11	AC014088	AC014088 Homo sapi
25	37.2	3.9	34214	16	ADRGENOME	U14448 Human adeno
26	37	3.8	446	13	G45875	g45875 24740_1 Zeb
27	36.8	3.8	310565	32	CBY62F5	AL [illegible] Caenorhab
c 28	36.8	3.8	4729	34	CBY59B7	CBY59B7 Caenorhabdi
29	36.8	3.8	34305	34	CEW02B12	Z66521 Caenorhabdi
c 30	36.8	3.8	155886	42	AC012134	AC012134 Homo sapi
31	36.6	3.8	132400	33	AC007742	AC007742 Homo sapi
32	36.6	3.8	7720	34	CEFAH11	M74777 P.falciparu
c 33	36.6	3.8	175793	40	AC011592	AC011592 Homo sapi
c 34	36.4	3.8	68033	32	PFMAL1P8	AL [illegible] Plasmodiu
c 35	36.4	3.8	2716	38	AF127287	AF127287 Dictyoste
c 36	36.2	3.7	235956	32	PFMAL4P2	AL [illegible] Plasmodiu
c 37	36.2	3.7	163606	42	AC011105	AC011105 Homo sapi
c 38	36	3.7	100000	10	AP000143	AP000143 Homo sapi
c 39	36	3.7	35091	10	AP000220	AP000220 Homo sapi
c 40	36	3.7	123681	11	HS22F11	AL [illegible] Homo sapi
41	36	3.7	110000	32	CEY102F5_0	AL [illegible] Caenorhab
42	36	3.7	35409	34	CEC44C11	Z84797 Caenorhabdi
c 43	35.8	3.7	116792	10	HS110H13	AL [illegible] Human DNA
44	35.8	3.7	151363	33	HSDJ493F7	AL [illegible] Homo sapi
45	35.8	3.7	32723	34	CEM88	Z84802 Caenorhabdi

ALIGNMENTS

RESULT 1
AB002365
LOCUS AB002365 5654 bp mRNA PRI 19-FEB-1999
DEFINITION Human mRNA for KIAA0367 gene, partial cds.
ACCESSION AB002365
VERSION AB002365.1 GI:2204674
KEYWORDS KIAA0367.
SOURCE Homo sapiens male brain cDNA to mRNA, clone:pBluescriptII SK plus clone:HH0141.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1. .5654
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .2151
                     /gene="KIAA0367"
     CDS             <1. .2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20822.1"
                     /db_xref="GI:2224675"
                     /translation="ALFEDSPELSTENPALVPEALLASDTCLDISEAAFDHSFSDASG
                     LNTSTGTIDDMSKLILSEGHPEIPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDLL
                     REPEHFLYGGDPPLEEDSLKQSLAPYIPPFDLSYLTEPAQSAETIEEAGSPEDESLGC
                     RAAFIVLSALPDRRSEGNQAFTKNRLPGSQIAVLHIREDPESVYLPVGAGSNILSPSN
                     VDWEVETTNEDLFAGGDIGFTNGASKEIPELEEEKTIFTKEPEQIKSEYKEERQTEKN
                     EDRHALHMDYILVNREENSHSKPETCEERESIAELELYVGSKETGLQGTQLASFPDTC
                     QPASLNERKGLSAEKMSSKGETRSSFESPAQDQSWMFLGHSEVSDFSLIARDSGPGWS
                     GKTVEPFSELGLGEGPQLQILEEMKPLESLALEEASGPVSQSQKSKSRGRAGPDAVIH
                     DSEWEMLSPQPVQKNMISDIEMEEETEFLELGTRISRPNGLLSEDVGMDIPFEEGVLS
                     PSAADMPPEPPNSLDLNDTHPRPIKLTAPNINLSLDQSEGSILSDDNLDSPDEIDINV
                     DELITPDEASFEYTGHFTANKDSGIFSESIPEYTAREEPEDNFLWRTVVIDEQEQF
                     IDMKVIEPYRRVISHGGYYGDGLNAIIVFAACFLPDSSFADYHYVMENLFLYVISTLE
                     LMVAEDYMIVYLNGATPKKEDAPAPLDEEMLPDD"
BASE COUNT     1564 a    1188 c    1198 g    1604 t
ORIGIN
```

```
  Query Match             93.4%;  Score 901.8;  DB 9;  Length 5654;
  Best Local Similarity   97.7%;  Pred. No. 1.2e-239;
  Matches  939;  Conservative    7;  Mismatches   12;  Indels   10;  Gaps    2;

Qy      1 atgaaaggaggaagacagaattccttgagctcggaaccaggatatcaagaccaaaatggacta 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1414 ATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGAACCAGGATATCAAGACCAAATGGACTA 1473

Qy     61 ctgtcagaggatgtaggaatggacatcccctttgaagagggcgtgctgagtcccagtgct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1474 CTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCAGTGCT 1533

Qy    121 gcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcctcggaga 180
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1534 GCAGACATGAGGCCTGAACCTCCTAATTCTCCTGGATCTTAATGACACTCATCCTCGGAGA 1593

Qy    181 atcaagctcacagccccaaaiatcaatctttctggaccaaagtgaaggatctattctc 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1594 ATCAAGCTCACAGCCCCAAAIATCAATCTTTCTGGACCAAAGTGAAGGATCTATTCTC 1653

Qy    241 tctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgataccc 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1654 TCTGATGATAACTTGGACAGTCCAGATGAAAT GACATCAATGTGGATGAACTTGATACC 1713
```

```
Qy    301 cccgatgaagcagattcttttgagtcccctggccatgatcccacagccaaaaaaiatcct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1714 CCCGATGAAGCAGATTCTTTTGAGTCACCTGGCCATGATCCCACAGCCAAAAAAIATCCT 1773

Qy    361 ggccaagagtcagagtctattccagctctacaggcctgaaaggggatttttgtttcctcct 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1774 GGCCAAGAGTCAGAGTCTATTCCAGCTCTACAGGCCTGAAAGGGGATTTTTGTTTCCTCCT 1833

Qy    421 ctttggatgacagtggtcattgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1834 CTTTGGAGGACAGTGGTCATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1893

Qy    481 cctctacaggagagtcatttctcaaaaaaaaattcaagatactctatgggaaagtctaaat 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1894 CCCTACAGGAGAGTCATTTCTCAAAAA--------AAGATACTCTATGGGAAAGTCTAAAT 1944

Qy    541 gccatcattgtgtttgccgcctgttttctgccagacagcagtcgtggcggattacgctgat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1945 GCCATCATTGTGTTTGCCGCCTGTTTTCTGCCAGACAGCAGTCGTGGCGGATTACGCTGAT 2004

Qy    601 gtcatggaaaaatcttttcctatatgtaatgaagtacttttagagttgatgatgaactaaaaaa 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2005 GTCATGGAAAAATCTTTTCCTATATGTAATGAAGTACTTTTAGAGTTGATGATGAACTAAAAAA 2064

Qy    661 tatatgattgtgtacttgaatggtgaaa-ccccaagaagaagatgccaaaaaaaaaaaaa 719
          ||||||||||||||||||||||||||| |||||||||| |||| |||||||||||||||
Db   2065 TATATGATTGTGTACTTGAATGGTGAAACCCCAAGAAGAAGATGCCAAAAAAAAAAAAA 2124

Qy    720 gatgaagaaaatgctaccagatgattgaggaaggtgaggaaaaaattaaaaatgttgaaa 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2125 GATGAAGAAATGCTACCAGATGATTGACGACGTTGAGGAAGAATTTAAAATTTTGAAA 2184

Qy    780 cattgttcatccatcttggttcatcagaagaatccttgcttgtgaacgagaattttataaa 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2185 CATTGTTCATCCATCTTGGTTCATCAGAAGAATCCTTGCTTGTGAACGAGAATTTTATAAA 2244

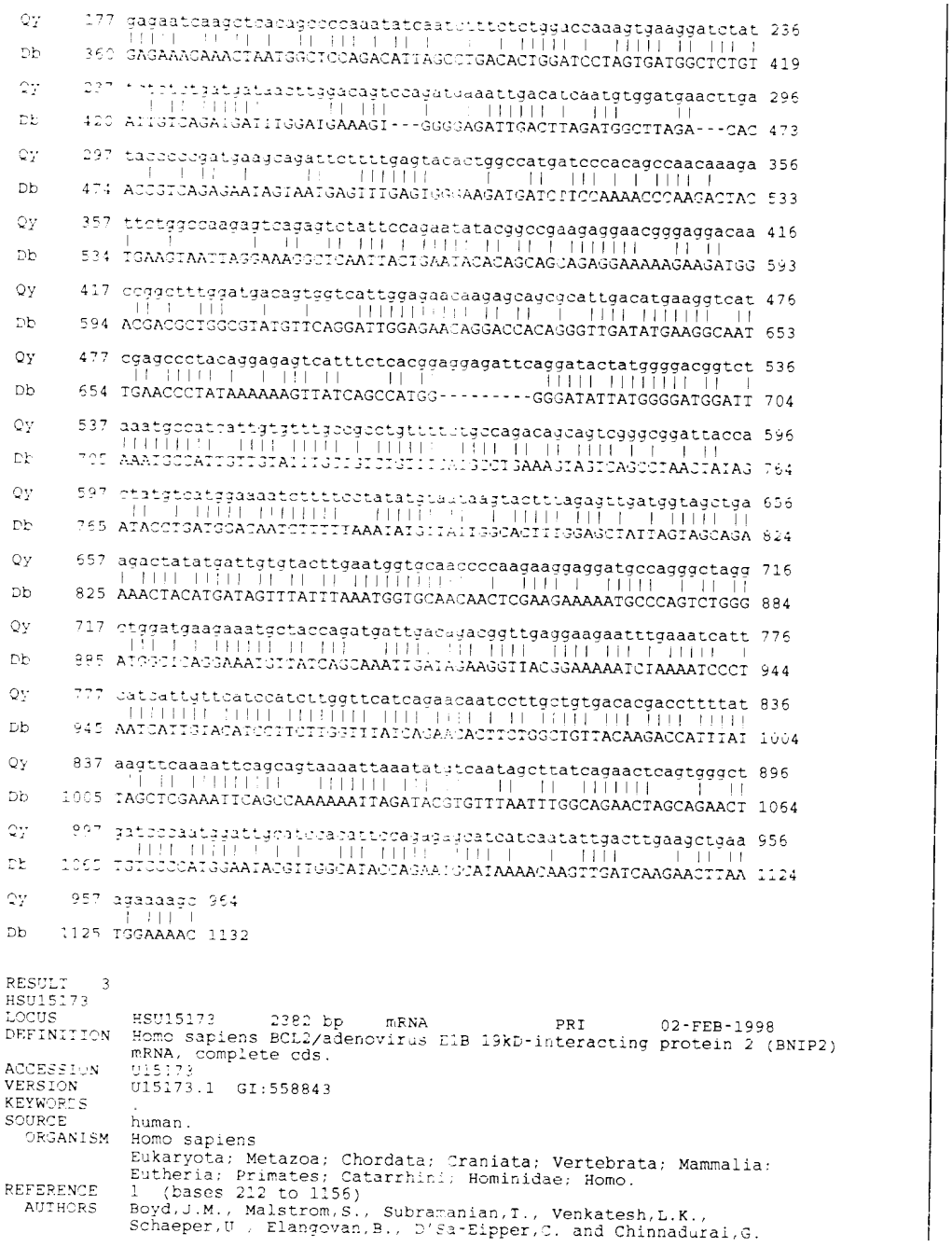
Qy    840 ttcaaaattcagcagtaaaaattaaatatgtcaatagcttatcagaagtcaattgaaattgat 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2245 TTCAAAATTCAGCAGTAAAATTAAATATGTCAATAGCTTATCAGAAGTCAATTGAAATTGAT 2304

Qy    900 cccaatggattgcatccacattcagaggaccatgatgaatattgaattttaaaaaaaaaaaa 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2305 CCCAATGGATTGCATCCACATTCAGAGGACCATGATGAATATTGAATTTTAAAAAAAAAAAA 2364

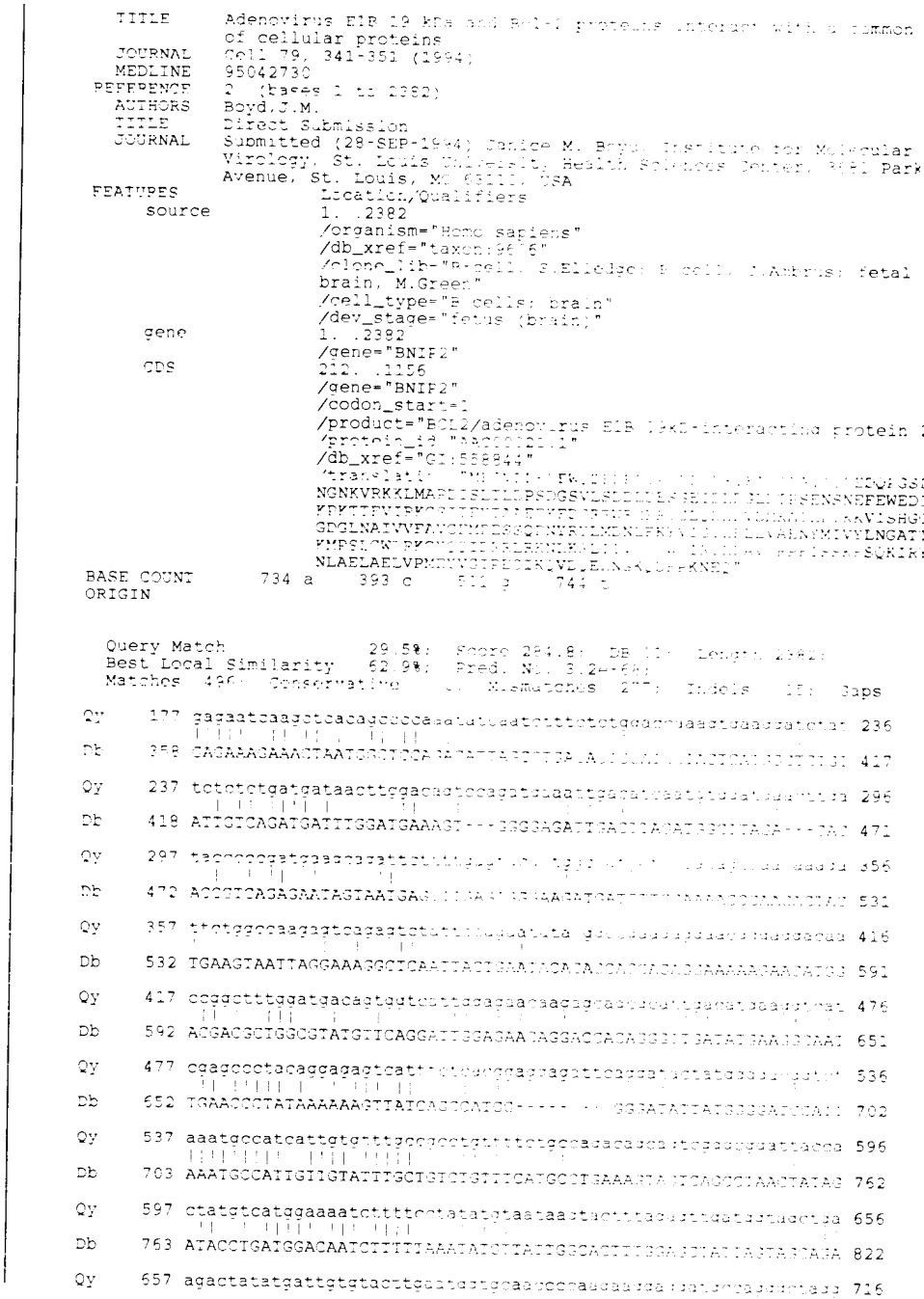
Qy    960 a 960
          |
Db   2365 A 2365
```

```
RESULT   2
AP028453
LOCUS       AP028453     1386 bp    DNA             PAT       [illegible]99
DEFINITION  Sequence 22 from patent US 5858678.
ACCESSION   AP028453
VERSION     AP028453.1  GI:5940426
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G.
  TITLE     Apoptosis-regulating proteins
  JOURNAL   Patent: US 5858678-A 22 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    336 g    373 t
ORIGIN
```

```
  Query Match             29.5%;  Score 284.8;  DB 5;  Length 1386;
  Best Local Similarity   62.9%;  Pred. No. 3e-68;
  Matches  496;  Conservative    0;  Mismatches  277;  Indels   15;  Gaps    3;
```



RESULT 3
HSU15173
LOCUS HSU15173 2382 bp mRNA PRI 02-FEB-1998
DEFINITION Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2) mRNA, complete cds.
ACCESSION U15173
VERSION U15173.1 GI:558843
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 212 to 1156)
AUTHORS Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K., Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.



TITLE Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set of cellular proteins
JOURNAL Cell 79, 341-351 (1994)
MEDLINE 95042730
REFERENCE 2 (bases 1 to 2382)
AUTHORS Boyd,J.M.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1994) Janice M. Boyd, Institute for Molecular Virology, St. Louis University Health Sciences Center, 3681 Park Avenue, St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1..2382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="B-cell, S.Elledge; B cell, J.Ambrus; fetal brain, M.Green"
/cell_type="B cells; brain"
/dev_stage="fetus (brain)"
gene 1..2382
/gene="BNIP2"
CDS 212..1156
/gene="BNIP2"
/codon_start=1
/product="BCL2/adenovirus E1B 19kD-interacting protein 2"
/protein_id="AAC50281.1"
/db_xref="GI:558844"
/translation="MHRTASHLGQNSPAMEDGSLEV
NGNKVRKKLMAPCISLILPSDGSVLHSLFSENSNEFEWEDDLP
KPKTTFVIRVAGSTTFRSATETLWEPWFEELTHPAVQLTVMFR
GDGLNAIVVFANCFMERGCPNSIFLMELEPKVVSGMLRVALNYMIVYLNGATTRR
PMPSLQWERKCWSSVRSLRRWLKELL
NLAELAELVPMDYYGSPSRIVLELNSALFKNEL"
BASE COUNT 734 a 393 c 511 g 744 t
ORIGIN

Query Match 29.5%; Score 284.8; DB 1; Length 2382;
Best Local Similarity 62.9%; Pred. No. 3.2e-60;
Matches 496; Conservative 0; Mismatches 277; Indels 15; Gaps 3;

```
Db    823 AAACTACATGATAGTTTATTTAAATGGTGCAATAACTCGAAGAAAAATGCCCAGTCTGGG 882

Qy    717 ctggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttgaaatcatt 776
          ||| | . |||||| || ||| ||||.|  | |||| |||| ||| | ||||| |
Db    883 AIGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAAAAATCTAAAATCCCT 942

Qy    777 catcattgttcatccatcttggttcatcagaaaaatccttgctgtgacacgacctttttat 836
          ||||| || |||||.|||||||| ||| ||  | | || ||||| ||| |||| ||||||
Db    943 AATCATTGTACATCCTTCTTGGTTTATCAGAAAACTTCTGGCTGTTACAAGACCATTTTAT 1002

Qy    837 aagttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaactcagtgggct 896
          || | |||| |||||  |||||| | | .   || |'  |||||||    |  ||
Db   1003 TAGCTCGAAATTCAGCCAAAAAATTAGATACGTGTTTAATTTGGCAGAACTAGCAGAACT 1062

Qy    897 gatcccaatggattgcatccacattccagagagcatcatcaatattgacttgaagctgaa 956
          ||'| ||||| | | |   ||| ||| |  ||| |   '  ||||     | || ||
Db   1063 TGTCCCCATGGAATACGTTGGCATACCAGAATGCATAAAACAAGTTGATCAAGAACTTAA 1122

Qy    957 agaaaagc 964
          | ||| '
Db   1123 TGGAAAAC 1130


RESULT   4
AF035207
LOCUS       AF035207     2213 bp    mRNA            ROD       25-FEB-1998
DEFINITION  Mus musculus Nip21 mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip21 gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, CuraGen Corporation,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1..2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit54 and D9Mit1"
     gene            1..2213
                     /gene="Nip21"
     CDS             134..1114
                     /gene="Nip21"
                     /note="associates with motifs of both the adenovirus E1B
                     19 kDa and Bcl-2 proteins"
                     /codon_start=1
                     /product="NIP21"
                     /protein_id="AAC04324.1"
                     /db_xref="GI:2911348"
                     /translation="MEGVPLQWQDEDPPIPLPEDDSIEADILDGTDPDPQPGSLEV
                     NGNKVRKKLMAPDISLTLDPGSLWSDDLDEAGEVDLEGLDTPSENSDEFEWEDDLP
                     KPKTIEVIPKGSITEYTAIEPKGDGPPWPMFRIGEQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESGQFNYRYLMDNLFKYVIGTLELLVAENYMIIYLNGAIIRP
                     KMPSLGWLPPCYQQIDRRLPKNLKSLIIVHPSWFIRTLLAVTRPFISSKFSQKIPYVF
                     NLAELAELVPMEYVGIPECIAAYEEEKFKKMQKRVDQELNGKQEPPKSEQ"
BASE COUNT      637 a    400 c    571 g    605 t
ORIGIN


  Query Match             99.3%;  Score: 2   ;  DB 12   Length 2213;
  Best Local Similarity   99.5%;  Pred. N   ; 9.9e-69;
```

```
  Matches  487;  Conservative    0;  Mismatches  265;  Indels   15;  Gaps    3;

Qy    178 agaatcaagctcacagccccaaatatgaatctttctctggaccaaagtgaaagtatttatt 237
          || |  || || |  |||||| | |||  ||  | ||||| |   |||||| ||
Db    281 AGGAAGAAACTGATGGCCCCAGACATCAGCCTGACCCTGGATCCTGGTGAAGATTTTCTG 340

Qy    238 ctctctgatgataacttggacagtccaagatgaaattgaacatcaatctgtaattgatttgat 297
          || ||||||   |||||     | |  ||  |||||  | | |      ||
Db    341 TGGTCCGATGAT---TTGGATGAAGCTGGAGAGGTTGACCTGGAAGGCTTAGA---TAT 394

Qy    298 accccgatgaagcagattcttttgattacactggccatgatcccacagccaacaaagat 357
          | | ||  |  |||  ||||||  |  || ||||  | |
Db    395 CCATCAGAGAACAGTGATGAGTTTGAGTGGGAAGATGATCTTCCCAAAAGCAAAAATTAT 454

Qy    358 tctggccaagagtcagagtctattccagaatatacggccgaagaagaacgtggaggtcaat 417
          | |      |  || || ||| | || || || ||| ||| ||  |  |  |
Db    455 GAAGTCATTAGGAAAGGCTCGATTACTGAGTACACAGCCACAGAAGAAAAAGGCGATGTA 514

Qy    418 cggctttggatgacagtggtcattgcagaacaagagcagccggcattgacatgaaggttgatt 477
          || |  |||  |  |  |||||||| || || || |  | | ||| | |
Db    515 CGCCGCTGGCGCATGTTCAGGATTGCAGAGCAGGACCACAGGGTGGACATCAAGGCGATT 574

Qy    478 gagccctacaggagagtcatttctccgggatggagatttcaggatactatgggggacatctca 537
          ||||||||||  | ||| |||   || ||          |||||| ||||||||||
Db    575 GAGCCCTACAAAAAAGTTATTAGCCATGG---------AGGATATTATGGTGACGGTTTA 625

Qy    538 aatgccatcattgtgtttgccgttttttttccccatgccttcttcccttttgtgtgtt 597
          |||||||  |||| ||||| | || |||  |||| || '|  ||| | |  | ||
Db    626 AATGCCATTGTIGTATTTGCTGTCCTGTTTCATCCCTGAGATCCCCATTTAACTATA 685

Qy    598 tatgtcatggaaaatcttttcctatatgtaataagtactttagaatttatttatgtgtt 657
          ||| | ||||| ||||| ||   || ||| ||  | ||||||||| | | |
Db    686 TATTTGATGGACAATCTCTTTAAATATGTTATTGGCACTTTAGAGCTTTTTTATTAGAT 745

Qy    658 gactatatgattgtgtacttgaatgtgtgaacgccaaggaaaaaaaatatgcaaaa taaaa 717
          |||| ||||  | ||||| |||||||| ||  |  |||| | | | | | |
Db    746 AACTACATGATCATTTACTTAAATGTTTGCAGAACTCGAAGAAAAAATGTTTAATTTTTA 805

Qy    718 tggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttgaaatcatt 777
          ||| | | || |||||||||||  || || ||||| | | | | | |  |
Db    806 TGGCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGGAAAAATGTTTAATTTTTA 865

Qy    778 atcattgttcatccatcttggttcatcagaaaaatccttgctgtgacacgacctttttat 837
          |||||||| || |||||||| || | | | | | | | | | | | | | | |
Db    866 ATCATTGTCACCCCTCTTGGTTTATCCGAACACTTCTGGCTGTTACAGCAATTTTAT 925

Qy    838 agttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaactcagtgggct 897
          |||||||| |||||  || | | ||| ||| | | | | | | | | | | |
Db    926 AGTTCAAAGTTCAGTCAGAAGATCAGATACGTCTTTAACTTGGCGCAATTTTTATATT 985

Qy    898 atcccaatggattgcatccacattccagagagcatcatcaatattga 944
          | || |||| |   ||  ||| |||||| ||||||    '  ||
Db    986 GTTCCCATGGAGTATGTCGGCATACCAGAGTGCATCAAACAGATATGA 1032


RESULT   5
AC009556
LOCUS       AC009556   169798 bp    DNA             HTG       15-NOV-1999
DEFINITION  Homo sapiens chromosome 10 clone RP11-57C21 map 10, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 10, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., [illegible], Morrison,M.,
```

```
            Baker,J., Baldwin,J., Barnes,N., Beckerly,P., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeAngelane,K., Depew,R.E., Dewar,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,J., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Kohoczky,J., Liew,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yo,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MII Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MII Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in 120 bases; agarose-fp
                Quality coverage: 4.1 in 120 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *              gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *              gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *              gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *              gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *              gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *              gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *              gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *              gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *              gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *              gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *              gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *              gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *              gap of unknown length
            *    37539    44831: contig of 7293 bp in length
            *              gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *              gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *              gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *              gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *              gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *              gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *              gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *              gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *              gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *              gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *              gap of unknown length
            *   140618   153842: contig of 13225 bp in length
            *              gap of unknown length
            *   153844   169798: contig of 15955 bp in length
  FEATURES             Location/Qualifiers
       source          1..169798
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="15"
                       /map="15"
                       /clone="RP11-57C21"
                       /clone_lib="RPCI-11 Human Male BAC"
  BASE COUNT    51330 a  34914 c  44407 g  39145 t
  ORIGIN


    Query Match             21.0%;  Score 214.4;  DB 44;  Length 169798;
    Best Local Similarity   99.5%;  Pred. No. 2.5e-47;
    Matches  203;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      136 gaaactcctaatttctctggatctaaaaa a tcatg tt.....t aggctaaaatt 195
            |||||||||||||||||||||||||||||| | ||||| |||||||||||||||||||
Db 119825 GAACCTCCTAATTCTCTGGATC.AAAAGA.AT.CA.CTTTAA.AAATAAGT.A.ATTT 119884

Qy      196 ccaaaatatcaatctttctctgc.....aaactcaggcgttattct.tttcatttcaa 255
            |||||||||||||||||||||||     ||||||||||||||||| | |||||||||||
Db 119885 CCAAAATATCAATCTTTCTCTGGACTAAAAT.GAAGGATTTATTCT.TTCATTTCAA 119944

Qy      256 gacaatccagatgaaattgacatcaatgtggatgaactgata.....tatat 315
            ||||||||||||||||||||||||||||||||||||||||     |||||
Db 119945 GACAATCCAGATGAAATTGACATCAATGT.GATGAACTGATA.AT.TTTTA.TAATTATAT 120004

Qy      316 tatttttgagtacactggccatgtt 449
            ||||||||||||||||||||||||||
Db 120005 TATTTTGAGTACACTGGCCATGTT 120128


RESULT   6
AC009556/c
LOCUS       AC009556   169798 bp    DNA             HTG       PRI 15-NOV-1999
DEFINITION  Homo sapiens chromosome 15, clone RP11-57C21, *** SEQUENCING IN
            PROGRESS *** 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479108
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
```

```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayne,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Rile,F., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap, version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in 120 bases; agarose-fp
                Quality coverage: 4.1 in 120 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *                   gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *                   gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *                   gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *                   gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *                   gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *                   gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *                   gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *                   gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *                   gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                   gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                   gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                   gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


  Query Match             17.9%;  Score 178.8;  DB 44;  Length 169798;
  Best Local Similarity   93.8%;  Pred. No. 3.8e-37;
  Matches  180;  Conservative    0;  Mismatches   12;  Indels     3;  Gaps    0;

Qy    335 atgatcccacagccaacaaagattctggccaagcagtcagagtctattccacaaaatacacc 394
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64014 AAGATCCCACAGCCAACAAAGATTCTGGCCAAGCAGTCAGAGTCTATTCCACAAAATACACC 63955

Qy    395 ccgaaagaggaacgggaggacaaxxxgxxxgaatgcagtggtcattgaaaaaaaggaaa 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63954 CCGAAGAGGAACCGGAGGACAAXXXGXXXTTGGAGGACAGTGGTCATTGAAAAACAAGAAA 63895

Qy    455 agcgcattgacatgaaggtcatgxxxxxxacaggagagtxxxxxxxxxxxxxxxxxxxxx 514
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63894 AGCGCATTGACATGAAGGTCATCXXXXXXACAGGAGAGTCATTCTCACTAAGTTAATT 63835

Qy    515 caggatactatg 526
          ||  |  |||
Db  63834 GAGCTCAGCATG 63823


RESULT   7
AC011488/c
LOCUS       AC011488   179301 bp    DNA             HTG       Jx-11-99
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_17[N]1, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *              gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *              gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *              gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *              gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *              gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *              gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *              gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *              gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *              gap of unknown length
            *     6972     7937: contig of 966 bp in length
            *              gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *              gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *              gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *              gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *              gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *              gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *              gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *              gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *              gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *              gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *              gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *              gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *              gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *              gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *              gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *              gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *              gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *              gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *              gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *              gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *              gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *              gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *              gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *              gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *              gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *              gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *              gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *              gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *              gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *              gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *              gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *              gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *              gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *              gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *              gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *              gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *              gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *              gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *              gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *              gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *              gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *              gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *              gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *              gap of unknown length
            *   112601   179301: contig of 66701 bp in length
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome "19"
                     /clone="CTD-2165K17;CTC13"
BASE COUNT    42023 a  47125 c  47[illegible] g  4259[illegible] t      [illegible] others
ORIGIN


  Query Match              9.5%;  Score 9[illegible];  DB 41;  Length 179[illegible];
  Best Local Similarity   65.7%;  Pred. No. 8[illegible]-15;
  Matches  134;  Conservative    [illegible];  Mismatches  [illegible];  Indels    [illegible];  Gaps    0;

Qy      134 ctgaaactcctaattctctggat[illegible]actatcct[illegible]  193

Db 124309 CAGCTCCTCCCAACACCCTAAA[illegible]  124247
```

```
Qy     194 ccccaaaatatcaatctttctctggaccaaagtaaaggatctattctctctgatgataact 253
           ||||| | |||||  |||||||||| || ||| | || ||  | || || |||||   ||
Db  124146 CCCCAAAATATCAATCATTTCTCTGGATCAAAGTAAAGGATCTATTCTCTCTGATGATAACT 124187

Qy     254 tggacagtccagatgaaattgacatcaatgtgaatgaacttgatacccccgatgaagcag 313
           |||| | || ||| | ||| | ||  || |||  | || ||||||||||  | |
Db  124188 TGGATAGCCCTGATAGCCCTGATATTAACCTTATGACATCCAGACCCCCGATGAGACCG 124127

Qy     314 attcttttgaagtacactggccatg 337
           | ||  | |||| |  ||  |||
Db  124126 ACTCGCTGGAGTTCCTGGGGAATG 124103


RESULT   8
AC015396/c
LOCUS       AC015396     78170 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015396
VERSION     AC015396.1  GI:6435939
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78170)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209602 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..78170
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21762 a   17118 c   17312 g   21978 t
ORIGIN


  Query Match             8.9%;  Score 85.6;  DB 43;  Length 78170;
  Best Local Similarity   60.5%;  Pred. No. 4.4e-13;
  Matches  161;  Conservative    0;  Mismatches   99;  Indels    6;  Gaps    1;

Qy     460 attgacatgaaagtcatcgagccctacaggatagtcatttctcacggaggagattcagga 519
           || ||||||  ||||||| |||||||||| | . ||  | || ||||| ||  ||
Db    8691 AICGACAIGCGGGICAIAGAGCCCIACAAGC!!GIGCIGTCCCACGGCGGCTAI------ 8637

Qy     520 tactatgggtacggtctaaatgccatcattg:tttgccgcctgttttctgccagacagc 579
           |  ' |  | ||||' || |||||  :  ||  |||||||   ||||| || ||
Db    8637 CICAAAICIGGCGGICAGAACGCCATIGICA:.IIIIGCGCCIGICACCIGCCGGATAGG 8578

Qy     580 agtcgaggcgaattaccactatgtcatggaaaatcttttcctatatgtaataagtacttta 639
           |||| || |||  ||| || ||||| | | || || || ||  | |  ||  |
Db    8577 ICGAGGGCGGACIACAGCIACGIGATGGACAA:.IIGTIICIGIACGIGGICAAGACCCIG 8518

Qy     640 tagttatatgatagctgaaaactatatgatt!:tacttgaatggtgcaacccccaagaagg 699
           |  | ||||  ||| |||||| ||      ||  || |||| |  ||  ||  |
Db    8517 CAACAGIIGGICACIGAIGACIAIGIGCIAA:.IAICIGCAIGGCGGCICCAATAGGCGI 8458

Qy     700 aggatgcaaaggctaggctggatgaa 725
           |  ||||  |  |||| | ||
Db    8457 AAIGIGCCACCIIIICCCIGGCTTAA 8432


RESULT   9
HSRHOGAPX
```

```
LOCUS       HSRHOGAPX    1430 bp    mRNA            PRI       25-AUG-1997
DEFINITION  H.sapiens rhoGAP protein.
ACCESSION   Z23024
VERSION     Z23024.1  GI:312211
KEYWORDS    rhoGAP protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   Hall,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1993) Alan Hall Dr., Cell and Molecular Biology,
            Institute of Cancer Research, Chester Beatty Laboratories, 239
            Fulham Road, London, SW3 6JB, England
FEATURES             Location/Qualifiers
     source          1..1430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="Male"
                     /tissue_type="fibrosarcoma"
                     /cell_line="HT1080"
     5'UTR           1..98
     CDS             99..1418
                     /codon_start=1
                     /product="rhoGAP protein"
                     /protein_id="CAA80560.1"
                     /db_xref="GI:312212"
                     /db_xref="SWISS-PROT:Q07960"
                     /translation="MDPLSELIDDLTLDDTSEALNLPLAPIDPPKWPSLEMPDFPKS
                     DDSKSSSPELVTHILKWDDPYYDIAPHQIVEVAGEDAYGRTIVRAADKMSPSHILDHS
                     FLLGYLKHTLDQTVESPMTLIYLHRGITSPNKPSLSWIRTAVSPPTRKPAKNIKALYI
                     VHPTMFIKTLLIIPKPLISPKPRQKIFYVNYLSFLSFHVKFFQHCIPKQVLKYDDFLK
                     STQYSPATAPKPMIPPPLPNLIPDVSIHIIPPNPPPPPPIVTPPITNAYLQAHALT
                     TESIFPPRSANTQVVPPVQQKYNMGLPVDFDQYNFLHIPAVIKTFLSELPEPLLIFDL
                     YPHVVGFLNIDESIEVPATLGVLQTLPDENYQVLSFLTAIIVQISAHSPLNKMTNTNL
                     AVVFGPNLLWAKDAAITLKAINPINTPTKFLLDHQGELPLSPIISSL"
BASE COUNT      332 a    453 c    365 g    280 t
ORIGIN


  Query Match             6.2%;  Score 60;  DB 9;  Length 1430;
  Best Local Similarity   48.3%;  Pred. No. 3.8e-06;
  Matches  168;  Conservative    0;  Mismatches  180;  Indels    1;  Gaps    0;

Qy     610 aatcttttcctatatgtaataagtacttttgaatttgatggtaacctaaaggctatatgatt 669
           || ||  |   ||  | |  |  | ||  |  |  |  |  |  |  |  |  |  |
Db     405 AAGCTCCTGGGGTACCTGAAGCACTACCCTGACCACTACTTTTTATACCCTTTTATACTT 464

Qy     670 gtgtacttgaatggtgcaacccccaagaaggatgccaaaactaaactaaatatatttttt 729
           ||||  || |   |      | ||    | | |||    ||  |||||  |
Db     465 CTGTATCTGCACCACGGCCTGACCAGCCACAACAAGCCCTCCCTCAGCTGTTTTTTTAA 524

Qy     730 tgctaccagatgattgacagacgg tgaggaagaaatttgaaatcattgatatatttttt 789
           |||||  |  ||||| |  ||  | |||||  |  |  |  |  |  |  |  |
Db     525 GCCTACCGGGAGTTTGACCGCAAATACAAAAAAAATATAAATTTTTTTATTTTTTTTT 584

Qy     790 ccatcttggttcatcagaacaatatttgttgtgacagacgcttttatagattcaattttt 849
           |||  |   |||||||| |||  | ||      |    ||  |  ||  |  |  |
Db     585 CCAACCATGTTCATCAAAACTCTCTTCATCCTTTCAACCCCTTTATTATTTTTAGTTT 644

Qy     850 agcagtaaaattaaatatgtcaatagcttatcagaactcagtgcgcttgatttcctatttt 909
           |    || ||    |||||  |  |  |  ||  |  |  |  |  |  |  |  |
Db     645 GGGCAGAAGATCTTCTATGTGAATTACCTGAGCGACCTGACTTAAATTTGTAATTTTT 704

Qy     910 tgcatccacattccagagagcatcatcgatattgacttgaaactaaa 955
```

```
Db    705 CAGCTGGGGATCCCTCGCCAAGTGCTCAAAATGACGACTTCCTGAAA 752


RESULT  10
HSU02570
LOCUS       HSU02570     3345 bp    mRNA            PRI       25-JAN-1994
DEFINITION  Human CDC42 GTPase-activating protein mRNA, partial cds.
ACCESSION   U02570
VERSION     U02570.1  GI:450780
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3345)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 268, 26059-26062 (1993)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 3345)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P., Self,A.J., Brill,S., van Erp,H.E.
            and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase Activating Protein for
            Rho-Related Small GTPases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3345)
  AUTHORS   Barfod,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1993) Elizabeth T. Barfod, Zeneca Inc., 440
            Point San Bruno Blvd., South San Francisco, CA 94080, USA
COMMENT     On Jan 27, 1994 this sequence version replaced gi:409026.
FEATURES             Location/Qualifiers
     source          1..3345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CMK"
     CDS             <1..1391
                     /codon_start=3
                     /function="GTPase activating protein for the rho family of
                     small GTP-binding proteins that is most effective against
                     CDC42"
                     /gene="CDC42 GTPase-activating protein"
                     /protein_id="AAA50042.1"
                     /db_xref="GI:409027"
                     /translation="DSSAGTSLLFPIPAFLSGTGLAMCPLSELIDELTDDTSEALN
                     QLKLASIDEKNWPSDEMPDPPKSDDSKSSSPELVTHLKWDDPYYDIARHQIVPVAGD
                     KVGPKITVFSACPMPPSHQ"HSKLLGYLKHTLDQYVESDYTLLYLHHGLTSDNKPSL
                     SWLRDAYREFDRKYKKNIKALYIVHPIMFIKTLLILFKPLISFKFGQKIFYVNYLSEL
                     SEHVPLELLCIPPQYLKYDFPLKSTQKSPATAPKPMPPRPPLPNQQFQVSLQHLIEFN
                     PEQEPIPIVLRETVAYLQAHALITEGIFRRSANIQVVREVQQKYNMGLPVDFDQYNEL
                     HLPAVILKIFLPELPEPLLTFDLYPHVVGFLNIDESQPVPATLQVLQTLPEENYQVLF
                     FLTAFLVQISAHSDQNKMINLNLAVVFGPNLLWAKDAAIILKAINPINTFTKFLLDHQ
                     GELFPSPDPSGL"
     misc_feature    753..795
                     /note="SH3-binding domain; capable of binding the SH3
                     domains of src and p85 alpha; aa residues 251-261"
     misc_feature    849..920
                     /note="active site box 1; aa residues 283-308"
     misc_feature    999..1091
                     /note="active site box 2; aa residues 333-363"
     misc_feature    1143..1277
                     /note="active site box 3; aa residues 383-425"
     polyA_signal    3328..3333
     polyA_site      3345
                     /note="38 A residues"
BASE COUNT      655 a   1061 c    855 g    774 t
ORIGIN


  Query Match             6.2%;  Score 6.:  DB 9;  Length 3345;
  Best Local Similarity   43.3%;  Pred. N.:  4 [illegible]
  Matches  169;  Conservative   [illegible]  Mismatches  [illegible]  Indels  [illegible]  Gaps  0;

Qy    610 aatottttcctatatgtaaataa[illegible]                                669

Db    378 AAGCTCCTGGGGTACCTGAAGCA[illegible]                                437

Qy    670 gtgtaattgaattgtgtaaaaa[illegible]                                729

Db    438 GTGTATCTGCACCACGGCCTGA[illegible]                                497

Qy    730 tgctaccagatgattgacagaa[illegible]                                789

Db    498 GCCTACCGGGGAGTTTCACCCCA[illegible]                               557

Qy    790 ccatcttagttcatcagaaacaat[illegible]                              849

Db    558 CCAACCATGTTCATCAAAACTC[illegible]                                617

Qy    850 agcagtaaaattaaatatgtcaa[illegible]                               909

Db    618 GGGCAGAAGATCTTTCTATGTGAA[illegible]                              677

Qy    910 tgcatccacattcagagagcat[illegible]                               957

Db    678 CACCTGGGCATCCCTCGCCAAGT[illegible]                               725


RESULT  11
AC012249/c
LOCUS       AC012249    58386 bp    DNA             HTG       [illegible]-99
DEFINITION  Homo sapiens clone [illegible] [illegible] WORKING DRAFT SEQUENCE.
ACCESSION   AC012249
VERSION     AC012249.1  GI:6091[illegible]
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone [illegible]
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Ball,L.J., Barna,N., [illegible], B., [illegible], B. Abbatter,B.,
            Brown,A., Castle,A., Chiantroli,M., Collins,S., Colman,M.A.,
            [illegible], DeArellano, [illegible], Doyle,M.,
            Ferreira,P., FitzHugh,W., Forest,C., Fraser,A., Harp,L.,
            Gelegan,J., Gerdyna,S., Grout,D., Hagos,B., Hamilton,A., Horton,L.,
            Howland,J.C., Johnson,P., Jones,D., Kamal,M., Kann,L., Klein,J.,
            [illegible], Lee,J., [illegible], Maddox,E., Mahlke,L., Matthews,N.,
            McEwan,P., McGurk,A., Mikkelsen,K., McLaughlin, [illegible], Mo,R.,
            Morrow,J., Naylor,J., Norman,C.H., [illegible], Ponnuraj,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Sanders,R., Shetty,P.,
            Stange-Thomann,N., Stavropoulos,N., Subramanian,A., Sutton,G.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., [illegible], Wortman,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Rolly,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-19[illegible], Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were [illegible] using RepeatMasker [illegible] &
            Green, P. (1996-199[illegible])
            http://[illegible]
            * NOTE: This record [illegible]
            * sequencing reads that have not been assembled [illegible]
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is [illegible]
            * arbitrary. Low pass sequence samples are useful for
            * identifying clones that may be [illegible]
            * overlap relationships among clones [illegible]
            * However, it should not be assumed that this clone
            * will be sequenced [illegible]
```

```
* the record is updated, the accession number will
* be preserved.
*        1      799: contig of 799 bp in length
*                   gap of unknown length
*      800     1536: contig of 737 bp in length
*                   gap of unknown length
*     1537     2265: contig of 729 bp in length
*                   gap of unknown length
*     2266     2972: contig of 707 bp in length
*                   gap of unknown length
*     2973     3725: contig of 753 bp in length
*                   gap of unknown length
*     3726     4475: contig of 750 bp in length
*                   gap of unknown length
*     4476     5247: contig of 772 bp in length
*                   gap of unknown length
*     5248     5992: contig of 745 bp in length
*                   gap of unknown length
*     5993     6730: contig of 738 bp in length
*                   gap of unknown length
*     6731     7474: contig of 744 bp in length
*                   gap of unknown length
*     7475     8210: contig of 736 bp in length
*                   gap of unknown length
*     8211     8928: contig of 718 bp in length
*                   gap of unknown length
*     8929     9640: contig of 712 bp in length
*                   gap of unknown length
*     9641    10387: contig of 747 bp in length
*                   gap of unknown length
*    10388    11120: contig of 733 bp in length
*                   gap of unknown length
*    11121    11879: contig of 759 bp in length
*                   gap of unknown length
*    11880    12648: contig of 769 bp in length
*                   gap of unknown length
*    12649    13441: contig of 793 bp in length
*                   gap of unknown length
*    13442    14200: contig of 759 bp in length
*                   gap of unknown length
*    14201    14936: contig of 736 bp in length
*                   gap of unknown length
*    14937    15720: contig of 784 bp in length
*                   gap of unknown length
*    15721    16404: contig of 684 bp in length
*                   gap of unknown length
*    16405    17121: contig of 717 bp in length
*                   gap of unknown length
*    17122    17847: contig of 726 bp in length
*                   gap of unknown length
*    17848    18601: contig of 754 bp in length
*                   gap of unknown length
*    18602    19316: contig of 715 bp in length
*                   gap of unknown length
*    19317    20058: contig of 742 bp in length
*                   gap of unknown length
*    20059    20787: contig of 729 bp in length
*                   gap of unknown length
*    20788    21518: contig of 731 bp in length
*                   gap of unknown length
*    21519    22254: contig of 736 bp in length
*                   gap of unknown length
*    22255    23015: contig of 761 bp in length
*                   gap of unknown length
*    23016    23737: contig of 722 bp in length
*                   gap of unknown length
*    23738    24453: contig of 716 bp in length
*                   gap of unknown length
*    24454    25192: contig of 739 bp in length
*                   gap of unknown length
*    25193    25937: contig of 745 bp in length
*                   gap of unknown length
*    25938    26683: contig of 746 bp in length
*                   gap of unknown length
*    26684    27431: contig of 748 bp in length
*                   gap of unknown length
*    27432    28179: contig of 748 bp in length
*                   gap of unknown length
*    28180    28941: contig of 762 bp in length
*                   gap of unknown length
*    28942    29667: contig of 726 bp in length
*                   gap of unknown length
*    29668    30386: contig of 719 bp in length
*                   gap of unknown length
*    30387    31105: contig of 719 bp in length
*                   gap of unknown length
*    31106    31889: contig of 784 bp in length
*                   gap of unknown length
*    31890    32643: contig of 754 bp in length
*                   gap of unknown length
*    32644    33422: contig of 779 bp in length
*                   gap of unknown length
*    33423    34169: contig of 747 bp in length
*                   gap of unknown length
*    34170    34901: contig of 732 bp in length
*                   gap of unknown length
*    34902    35643: contig of 742 bp in length
*                   gap of unknown length
*    35644    36403: contig of 760 bp in length
*                   gap of unknown length
*    36404    37119: contig of 716 bp in length
*                   gap of unknown length
*    37120    37831: contig of 712 bp in length
*                   gap of unknown length
*    37832    38555: contig of 724 bp in length
*                   gap of unknown length
*    38556    39332: contig of 777 bp in length
*                   gap of unknown length
*    39333    40130: contig of 798 bp in length
*                   gap of unknown length
*    40131    40891: contig of 761 bp in length
*                   gap of unknown length
*    40892    41629: contig of 738 bp in length
*                   gap of unknown length
*    41630    42372: contig of 743 bp in length
*                   gap of unknown length
*    42373    43156: contig of 784 bp in length
*                   gap of unknown length
*    43157    43889: contig of 733 bp in length
*                   gap of unknown length
*    43890    44610: contig of 721 bp in length
*                   gap of unknown length
*    44611    45411: contig of 801 bp in length
*                   gap of unknown length
*    45412    46177: contig of 766 bp in length
*                   gap of unknown length
*    46178    46940: contig of 763 bp in length
*                   gap of unknown length
*    46941    47711: contig of 771 bp in length
*                   gap of unknown length
*    47712    48462: contig of 751 bp in length
*                   gap of unknown length
*    48463    49258: contig of 796 bp in length
*                   gap of unknown length
*    49259    50004: contig of 746 bp in length
*                   gap of unknown length
*    50005    50781: contig of 777 bp in length
*                   gap of unknown length
*    50782    51487: contig of 706 bp in length
*                   gap of unknown length
*    51488    52199: contig of 712 bp in length
*                   gap of unknown length
*    52200    52928: contig of 729 bp in length
*                   gap of unknown length
*    52929    53736: contig of 808 bp in length
*                   gap of unknown length
```

```
            *    53737    54547: contig of 811 bp in length
            *           gap of unknown length
            *    54548    55305: contig of 758 bp in length
            *           gap of unknown length
            *    55306    56070: contig of 765 bp in length
            *           gap of unknown length
            *    56071    56822: contig of 752 bp in length
            *           gap of unknown length
            *    56823    57607: contig of 785 bp in length
            *           gap of unknown length
            *    57608    58386: contig of 779 bp in length.

  Query Match              5.8%;  Score 56;  DB 42;  Length 58386;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-05;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggaggaggagacagagttccttgagctcgaaccaggatatcaagaccaaatgg 56
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7672 ATGGAGGAGGAGACAGAGTTCCTTGAGCTCGAACCAGGATATCAAGACCAAATGG 7617


RESULT  12
AC014963
LOCUS       AC014963   126121 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014963
VERSION     AC014963.1  GI:6436372
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 126121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211484 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    36005 a  25605 c  26872 g  37639 t
ORIGIN


  Query Match              5.6%;  Score 54;  DB 43;  Length 126121;
  Best Local Similarity   48.4%;  Pred. No. 0.00024;
  Matches  150;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    1;

Qy    647 tggtagctgaagactatatgattgtgtacttaaatggtgcaaccccaagaaggaggatgc 706
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17342 TTGTGGAGAACGACTACATTTTGGTTTACTTTCATCAAGGCTTCAAGGAGGACAACAAAC 17401

Qy    707 cagggctaagctggatgaagaaatgctacctatgattgacagacggttgaggaagaatt 766
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17402 CATCCGCACAGTTCCTGTGGAATTCGTACAAAGAGCTGGATCGCAACTTCCGCAAGAATC 17461

Qy    767 tgaaaatcattcatcattgttcatcctttgttcatcagaacaatcctttgtgtgacac 826
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17462 TGAAGACGTTGTATGTGGTTCATCCGACGTGCTTCATACGGGTCATTTGGAACTTCTTTA 17521

Qy    827 gagcttttataagttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaac 886
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  17522 GCCCGTTCATTAGCGACAAGTTCGGCAAGAACTAGTCTACATCTCGAGCTTGGACGAGT 17581
```

```
Qy    887 tcaatgagctgatcctaaatggattttatgttgaaaattgtaa 946
          |||||||||||||||||||||||||||||||||||||||||||||
Db  17582 TGTGCCAGGCGCTCGGCCTGAATAATCTGAATCTGTGATAATATCTTTATTTGAAT 17641

Qy    947 tgaagctgaa 956
          ||||||||||
Db  17642 ACAAACTGAA 17651


RESULT  13
HSU62794
LOCUS       HSU62794     1378 bp    mRNA            PRI       11-AUG-1997
DEFINITION  Human CDC42 GAP-related protein mRNA, partial cds.
ACCESSION   U62794
VERSION     U62794.1  GI:2216170
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 268 (35), 26350-26353 (1993)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP, a GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   3  (bases 1 to 1378)
  AUTHORS   Barfod,E.T. and Ashkenazi,A.
  TITLE     An effector function for CDC42GAP that modulates microtubule
            spindle
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Chan,M., Kuang,W.J. and Ashkenazi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1996) Molecular Oncology, Bristol-Myers Squibb Co., 4
            San Bruno Blvd., S. San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1378
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="JM"
                     /cell_type="primary cyte"
     CDS             <1..672
                     /note="CDC42-GAP-related protein encoded by GenBank Accession Number U62794;
                     insertion in the rhoGAP domain with homology to site
                     boxes 2 and 3 out of frame and untranslated"
                     /codon_start=1
                     /product="CDC42 GAP-related protein"
                     /protein_id="AAB62794.1"
                     /db_xref="GI:2216171"
                     /translation="LNPLFSHSSLGALSVPPYSMNRLSVNAPITRLLSCPFCL
                     CRYFFNTPALYCHMTRVLLLLPNLLSYRNQLPIPPVNVLLQLGI
                     PPQVLKTDDFLKSLKSLATNKFMPRFSTMSLVPLPKNILEPIPIVLR
                     ETVAVLQAHALTTRIPRPSANTCNVREVLLRYSMLTTPSRTPRLLSR*PHPRAACG
                     FRPVQ"
     misc_feature    370..402
                     /note="encodes SH3-binding domain; partial; 3' binding src
                     and p85"
     misc_feature    465..543
                     /note="active site box 1"
     misc_feature    595..640
                     /note="55 nucleotide insertion into rhoGAP domain; critical
                     with TTC/GTT; insertion between codons 1 and 2
                     shifts translational frame such that a stop codon occurs
```

```
                     at 670, and active site boxes 2 and 3 are not made"
     misc_feature    671. .763
                     /note="out-of-frame active site box 2"
     misc_feature    821. .949
                     /note="out-of-frame active site box 3"
     polyA_site      1378
                     /note="21 A nucleotides"
BASE COUNT      288 a    468 c    328 g    294 t
ORIGIN


  Query Match             4.6%;  Score 44.6;  DB 10;  Length 1378;
  Best Local Similarity   50.7%;  Pred. No. 0.068;
  Matches  107;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy     747 cagacggattgaggaagaatttgaaatcattcatcattgttcatccatcttggttcatcag 806
          |||  ||  | ||||||  | ||  | ||   ||| || |||||| |   ||||||||
Db     132 CTGCAGGTACAAGAAGAACATCAAGGCCTTGTACATCGTGCATCCAACCATGTTCATCAA 191

Qy     807 aacaatcattgctgtgacacgaccttttataattcaaaattcagcagtaaaattaaata 866
           |||  | ||      |          || | || | |  | || ||| |     || ||    ||
Db     192 AACTCTGCTCATCCTCTTCAAGCCCCTCATCAGCTTCAAGTTCGGGCAGAAGATCTTCTA 251

Qy     867 tgtcaatagcttatcagaactcagtgggctgatccaatggattgcatccacattccaga 926
           ||| ||  | |   || || | | |          |||'     |    || ||
Db     252 TGTGAATTACCTGAGCGAGCTGAGCGAGCACTTGAAGCTGGAGCAGCTGGGGATCCCTCG 311

Qy     927 aagcatcatcaatattgacttgaagctgaaa 957
                 |  ||||   |'||       ||| ||
Db     312 CCAAGTTCTCAAATATTGATGATTTCCTGAAA 342


RESULT  14
I66494/c
LOCUS       I66494          7,1.6 bp    DNA             PAT       29-SEP-1997
DEFINITION  Sequence 14 from patent US 567 367
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Schellinger,F. and Falkner,F.Guenter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             4.4%;  Score 42.6;  DB 5;  Length 7218;
  Best Local Similarity   5.1%;  Pred. No. 0.28;
  Matches   21;  Conservative  214;  Mismatches  178;  Indels    0;  Gaps    0;

Qy     358 totggccaagagtcaggagtctattccagaatatacggccgaaggaggaacgggaggacaac 417
           ||| ''||     |    ||  ||||'     |  :.....  .  .  .  .  .  .
Db    1472 TCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRR 1413

Qy     418 cggctttggattacagtggtcattggagaacagagcagcgcaattgacatgaaggtcatc 477
           ::    ::: :: :: :: :  : ..... ... .. .. .. .. .. .  .  .  .
Db    1412 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1353

Qy     478 gagccttacagagagtcattttctcaccgagtgattcaggctactatgggacggtcta 537
           ::    : ::::::  .  .........  ...  . : :::: ::  .  .  .
Db    1352 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1293

Qy     538 aatgccatcattgtgtttgccgcctgtttttcrccagacagcagtcgggcggattaccac 597
           :: : :  : :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    1292 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1233

Qy     598 tatgtcatggaaaaatctttcctatatgtaatgagtactttaaaatttttttt 657
           : :  : :::::   :  : :  :: ::: :   :::: :: :: ::  :::
Db    1232 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1173

Qy     658 gactatatgattgtgtacttgaatggtcaaccccaagaaggaaatgccacttttaaaa 717
           .  .  :  :  .  .  .  .  . .. . .  .  .  .  .  .  .  .  .  .
Db    1172 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1113

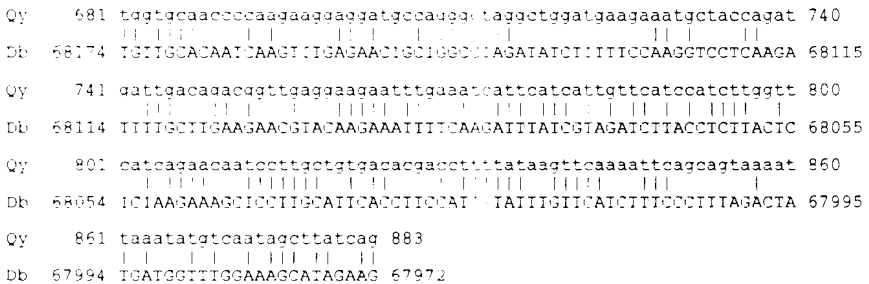
Qy     718 tggatgaagaaatgctaccagatgattgacagacggttgaggaaagaatttgaa 773
           .  .  .  .  .  .  .  .  .  .  .  .  :  :::::::  :
Db    1112 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPATGTAA 1062


RESULT  15
HS51J23/c
LOCUS       HS51J23     86004 bp    DNA             PRI       1-N-1999
DEFINITION  Human DNA sequence from clone 51J23 on chromosome X q26.3-27.3.
            Contains an EST and GSSs, complete sequence.
ACCESSION   AL031312
VERSION     AL031312.1  GI:3550064
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86004)
  AUTHORS   Wilkinson,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries, humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 8, 1998 this sequence version replaced gi:3445454.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 51J23. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat or gene elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre Chromosome X
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            51J23 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .86004
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q26.3 27.3"
                     /clone="RP1-51J23"
                     /clone_lib="RPCI-1"
     repeat_region   1200. .1556
                     /note="MER7 repeat matches 8. .335 of consensus"
     repeat_region   2508. .2531
                     /note="12 copies 2 mer at 96% conserved"
     repeat_region   2586. .2637
                     /note="13 copies 4 mer, tato 90% conserved"
     repeat_region   2654. .3564
                     /note="L1 repeat matches 4430. .5386 of consensus"
     repeat_region   3417. .3930
                     /note="L1MA3 repeat matches 1. .514 of consensus"
     repeat_region   3927. .4451
                     /note="L1MB7 repeat matches 777. 6111 of consensus"
     repeat_region   4449. .4713
```

```
                     /note="L1 repeat. matches 3469. .3746 of consensus"
     repeat_region   4712. .4992
                     /note="L1MB8 repeat. matches 649  920 of consensus"
     repeat_region   6899. .7305
                     /note="MSTA repeat. matches 3. .426 of consensus"
     repeat_region   7306. .8961
                     /note="MST-INTERNAL repeat. matches 1. .1639 of consensus"
     repeat_region   8974. .9382
                     /note="MSTA repeat. matches 1. .426 of consensus"
     repeat_region   9383. .9422
                     /note="20 copies 2 mer ca 85% conserved"
     repeat_region   9967. .10029
                     /note="L1MA6 repeat. matches 420. .482 of consensus"
     repeat_region   10365. .10565
                     /note="L1ME2 repeat. matches 512. .727 of consensus"
     repeat_region   10445. .10538
                     /note="L1ME3A repeat. matches 609  698 of consensus"
     repeat_region   13436. .13797
                     /note="MLT1A1 repeat. matches 1  365 of consensus"
     repeat_region   13822. .14162
                     /note="MLT1-INTERNAL repeat. matches 19  393 of consensus"
     repeat_region   14153. .14482
                     /note="MLT1-INTERNAL repeat. matches 427. .738 of
                     consensus"
     repeat_region   14461. .14529
                     /note="MLT1-INTERNAL repeat. matches 807. .877 of
                     consensus"
     repeat_region   14911. .15127
                     /note="MLT1-INTERNAL repeat. matches 942. .1165 of
                     consensus"
     repeat_region   complement(17602. .17902)
                     /note="AluYa5 repeat. matches 301  .1 of consensus"
     repeat_region   18470. .18562
                     /note="MIR repeat. matches 49  141 of consensus"
     repeat_region   complement(18701. .18989)
                     /note="AluSx repeat. matches 301. .13 of consensus"
     repeat_region   20736. .20787
                     /note="26 copies 2 mer ag 94% conserved"
     repeat_region   20743. .20786
                     /note="11 copies 4 mer gaga 100% conserved"
     repeat_region   21576. .21829
                     /note="MIR repeat. matches 2.  262 of consensus"
     repeat_region   complement(24250. .24501)
                     /note="L1 repeat. matches 5142. .4892 of consensus"
     repeat_region   25480. .26515
                     /note="L1 repeat. matches 4354  5390 of consensus"
     repeat_region   26954. .27251
                     /note="L1PB2 repeat. matches 1  902 of consensus"
     repeat_region   27502. .31652
                     /note="L1 repeat. matches 1235  5379 of consensus"
     repeat_region   31515. .31724
                     /note="L1PA9 repeat. matches 1. .226 of consensus"
     repeat_region   complement(31724  32008)
                     /note="L1PA6 repeat. matches 888  605 of consensus"
     repeat_region   32009. .32701
                     /note="L1PA15 repeat. matches 208.  892 of consensus"
     repeat_region   33027. .33330
                     /note="AluSx repeat. matches 1. .300 of consensus"
     repeat_region   complement(33469. .33965)
                     /note="MLT2_Internal repeat. matches 3981. .3786 of
                     consensus"
     repeat_region   complement(33998. .34301)
                     /note="MLT2_Internal repeat. matches 4303. .3486 of
                     consensus"
     repeat_region   34847. .34933
                     /note="AluSx repeat. matches 170. .256 of consensus"
     repeat_region   complement(34967. .35275)
                     /note="MLT2_Internal repeat. matches 3450  3130 of
                     consensus"
     repeat_region   36912. .37069
                     /note="MIR repeat. matches 20. .166 of consensus"
     repeat_region   37677. .37741
                     /note="MIR2 repeat. matches 82. .146 of consensus"
```

```
     repeat_region   39021. .39212
                     /note="MER2B repeat. matches [illegible] of consensus"
     repeat_region   39308. .39721
                     /note="MER2B repeat. matches [illegible] of consensus"
     repeat_region   39717. .40647
                     /note="MER2B repeat. matches [illegible] of consensus"
     repeat_region   40102. .40202
                     /note="L1 [illegible] of consensus"
     repeat_region   41104. .41396
                     /note="AluSg repeat. matches 1. .296 of consensus"
     repeat_region   41472. .44726
                     /note="L1 repeat. matches 2127. .734 of consensus"
     repeat_region   44502. .44524
                     /note="L1PB2 repeat. matches 1. .69 of consensus"
     repeat_region   47699. .47704
                     /note="9 copies 4 mer tgaa 88% conserved"
     repeat_region   complement(47699. .48079)
                     /note="L1PA2 repeat. matches 6[illegible] of consensus"
     repeat_region   complement(48774. .48980)
                     /note="L1 repeat. matches 534. .1124 of consensus"
     repeat_region   complement(49454. .50 10)
                     /note="MIR2 repeat. matches [illegible] of consensus"
     repeat_region   complement(50 32. .50 163)
                     /note="MIR repeat. matches 2[illegible] of consensus"
     repeat_region   51108. .51272
                     /note="MIR2 repeat. matches 1. .148 of consensus"
     repeat_region   51560. .51744
                     /note="MIR repeat. matches 1[illegible] of consensus"
     repeat_region   52901. .52940
                     /note="10 copies 4 mer aaaa 90% conserved"
     repeat_region   52901. .52996
                     /note="19 copies 2 mer ag 89% conserved"
     repeat_region   53268. .53387
                     /note="MIR2 repeat. matches 14. .142 of consensus"
     repeat_region   53533. .53652
                     /note="2 copies 57 mer 96% conserved"
     repeat_region   58593. .58612
                     /note="20 copies 2 mer ag 80% conserved"
     prim_transcript complement(59[illegible]. .59124)
                     /note="match: EST AA[illegible]25.3"
     repeat_region   58965. .59014
                     /note="L1PA11 repeat. matches 84[illegible] of consensus"
     repeat_region   59961. .60276
                     /note="MLT1A1 repeat. matches 1. .317 of consensus"
     repeat_region   complement(6[illegible]7. .6[illegible]59)
                     /note="MLT1A2 repeat. matches 17[illegible] of consensus"
     repeat_region   62307. .62677
                     /note="AluY repeat. matches 1. .[illegible] of consensus"
     repeat_region   62678. .62744
                     /note="21 copies 2 mer ag 90% conserved"
     repeat_region   complement(6[illegible]1. .6[illegible]69)
                     /note="MIR2 repeat. matches 1[illegible] of consensus"
     misc_feature    complement(6[illegible]3. .6[illegible]06)
                     /note="match: EST AA[illegible]44[illegible]"
     repeat_region   65813. .66[illegible]
                     /note="AluS repeat. matches 1. .[illegible] of consensus"
     repeat_region   complement(6[illegible]. .6[illegible]08)
                     /note="MLT1B repeat. matches [illegible] of consensus"
     repeat_region   complement(6[illegible]. .6[illegible]50)
                     /note="MLT1B repeat. matches [illegible] of consensus"
     repeat_region   complement(6[illegible]. .6[illegible]38)
                     /note="MER2A repeat. matches [illegible] of consensus"
     misc_feature    complement(6[illegible]. .6[illegible]79)
                     /note="match: EST AA[illegible]91"
     repeat_region   complement(6[illegible]2. .6[illegible]30)
                     /note="AluSx repeat. matches [illegible] of consensus"
     repeat_region   69017. .69[illegible]
                     /note="AluSq repeat. matches [illegible] of consensus"
     repeat_region   complement(6[illegible]. .69744)

Query Match              4.3%;  Score 41.4;  DB [illegible]  length [illegible]  4
Best Local Similarity    [illegible]%;  Pred. No. [illegible]
Matches  [illegible];  Conservative  [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  [illegible]
```



Search completed. February 9, 2000, 07.59.25
Job time: 12879 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on. February 9, 2000, 07:59:25 ; Search time 5888.89 Seconds
(without alignments)
-1529.306 Million cell updates/sec

Title: US-09-174-937-7
Perfect score: 2966
Sequence: 1 gtcgacccacgcgtccgggc.....aaaaaaaaaagggcggccgc 2966

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold . 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2701.4	91.1	169798	44	AC009556	AC009556 Homo sapi
2	562	18.9	58386	42	AC012249	AC012249 Homo sapi
c 3	426.2	14.4	58386	42	AC012249	AC012249 Homo sapi
4	282.4	9.5	5054	9	AB028965	AB028965 Human mRN
c 5	106.8	3.6	169798	44	AC009556	AC009556 Homo sapi
c 6	100.6	3.4	179301	41	AC011489	AC011489 Homo sapi
c 7	100.4	3.4	158282	38	HS730H16	AL122[illegible] Homo sapi
c 8	96.4	2.9	22242	34	PFMAL8P1	L47[illegible] Plasmodium
c 9	85.4	2.9	314303	41	AC006279	AC006279 Plasmodiu
10	84	2.8	94387	40	AC011716	AC011716 Homo sapi
11	83.6	2.8	237090	33	AC005077	AC005077 Homo sapi
12	83	2.8	179027	44	AC011229	AC011229 Homo sapi
c 13	82.8	2.8	103344	10	HS1100E15	AL035551 Human DNA
14	81.2	2.7	225956	32	PFMAL4P1	AL035475 Plasmodiu
15	81.2	2.7	194355	33	AL133216	AL133216 Homo sapi
16	81	2.7	103344	10	HS1100E15	AL035551 Human DNA
c 17	80.6	2.7	206784	11	HS932I7	AL022411 Homo sapi
c 18	80.6	2.7	110000	32	PFMAL4P1_2	Continuation (3 of
c 19	80.2	2.7	204027	42	AC007159	AC007159 Homo sapi
c 20	80	2.7	144107	32	CNS1D0T8	AL132666 Homo sapi
21	79.8	2.7	940	9	CNS01B0V	AL118078 Botrytis
c 22	79.6	2.7	108908	34	PFMAL8P1	AL035[illegible] Plasmodiu
23	79	2.7	170125	38	AC007465	AC007465 Homo sapi
24	79	2.7	184553	43	AC013452	AC013452 Homo sapi
25	78.6	2.7	190201	40	AC007277	AC007277 Homo sapi
c 26	78.6	2.7	169792	44	AC013567	AC013567 Homo sapi
27	78.4	2.6	696	8	CNS011EC	AL118078 Botrytis
28	78.2	2.6	204027	42	AC007159	AC007159 Homo sapi
29	78	2.6	206784	11	HS932I7	AL022411 Homo sapi
30	78	2.6	253807	34	PFMAL8P1	AL034556 Plasmodiu
31	78	2.6	14867	35	AE001398	AE001398 Plasmodiu
c 32	78	2.6	43427	41	AC006[illegible]	AC006[illegible] Homo sapi
c 33	78	2.6	184553	43	AC013452	AC013452 Homo sapi
34	77.6	2.5	158962	33	AC007679	AC007679 Homo sapi
35	77	2.6	162075	10	HS12T03	AL[illegible] Homo sapi
36	77	2.6	174427	32	PFMAL1P4	AL[illegible] Plasmodiu
c 37	77	2.6	144358	42	AC010811	AC010[illegible] Homo sapi
38	76.4	2.6	91017	11	HS86D21	AL022344 Human DNA
c 39	76.4	2.6	110000	32	PFMAL3P1_1	Continuation (2 of
40	76.4	2.6	159066	40	AC[illegible]41	AC[illegible]41 Homo sapi
c 41	76	2.6	308703	32	HS1[illegible]A22	AL[illegible]72 Homo sapi
42	75.8	2.6	166977	40	AC007[illegible]	AC007[illegible] Homo sapi
43	75.8	2.6	160322	44	AC011[illegible]	AC011[illegible] Homo sapi
44	75.8	2.6	154347	44	AC011[illegible]	AC011[illegible] Homo sapi
45	75.4	2.5	74586	11	HS2[illegible]E15	AL[illegible]82 Human DNA

ALIGNMENTS

RESULT 1
AC009556
LOCUS AC009556 169798 bp DNA HTG [illegible]-NOV-1999
DEFINITION Homo sapiens chromosome 19 clone [illegible] map 19, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC009556
VERSION AC009556.3 GI:6479158
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,T., Baldwin,J., Barna,N., Beckerly,P., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depaple,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Levine,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,T., Nilliff,M., O'Leary,T., O'Connell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torrealba-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute / MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.genome.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156043 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases, agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence.  It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *                   gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *     8009     9287: contig of 1279 bp in length
            *                   gap of unknown length
            *     9288    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *                   gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *                   gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *                   gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *                   gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *                   gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *                   gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *                   gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                   gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                   gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                   gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *   153844   169798: contig of 15955 bp in length
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34467 g  49065 t      2 others
ORIGIN


  Query Match             91.1%;  Score 2701.4;  DB 44;  Length 169798;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2749;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    4;

Qy     203 gaacctcctaattctctggatcttaatgacactcatcctcggaaaatcaaatctacaaat 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119825 GAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAAAATCAAATCTACAAAT 119884

Qy     263 ccaaaatatcaatctttctctggaccaaaatgaaggatctattctttctgaaaaaacctct 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119885 CCAAAATATCAATCTTTCTCTGGACCAAAATGAAGGATCTATTCTTTCTGAAAAAACCTCT 119944

Qy     323 gaagatcagaatgaaattgacataaatgtgaatgaattgaatgaaaaaatgaaaaaaaaa 382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119945 GAGAGTCCAGATGAAATTGACATAAATGTGAATGAATTGAATGAAAAAATGAAAAAAAAA 120004

Qy     383 tcttttgagtacactggccatgctaatcctagtttggcaaaaccaaaatttattatta 442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  120005 TCTTTTGAGTACACTGGCCATGCTAATCCTAGTTTGGCAAAACCAAAATTTATTATTA 120064

Qy     443 aataagtaaaagatttctaacagaaattgattttgtgccaatgaatgtttttgtgtct 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  120065 AATAAGTAAAAGATTTCTAACAGAAATTGATTTTGTGCCAATGAATGTTTTTGTGTCT 120124

Qy     503 ttctagaagcttctgttttattaaaattgtaggtggtatgtattgtatttgtgtct 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  120125 TTTCTAGAAGCTTCTGTTTTTATAAAATTGTAGGTGGTATGTATTGTATTTGTGTCT 120184
```



```
Db 122341 AACAATGAAGTACAGAGAAACTTAACTAATTTTTGTAAGGTCACACAGTTCATTGTTTGCCC 122400

Qy   2782 ttagttcctttccatgctgcttccagtgaatatgctaacaatgaatgggaaaagggtct 2842
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122401 TTAGTTCCTTGCCCATGCTGCTTCCAGTGAATATGCTAACAATGAATGGGAAAAGGGTCT 122460

Qy   2843 gttaccagattggaacttacatacttgaaagccaacatgatagctcttgctcacatttagg 2902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122461 GTTACCAGATTGGAACTTACATACTTGAAAGCCAACATGATAGCTCTTGCTCACATTTAGG 122520

Qy   2903 cctcttttttatccagttttcctgacttgcgggagactctaaaaaaaaaaaaaaaaggggcg 2961
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||   |||
Db 122521 CCTCTTTTTTATCCAGTTTTCCTGACTTGCGGGAGACTCTAAAAAAAAAAAAAAGTAAGCG 122579


RESULT   2
AC012249
LOCUS       AC012249    58386 bp    DNA             HTG       21-OCT-1999
DEFINITION  Homo sapiens clone 16_N_23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC012249
VERSION     AC012249.1  GI:6091795
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone 16_N_23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerle,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,E., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vasiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This record contains 78 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      799: contig of 799 bp in length
            *                    gap of unknown length
            *      800     1536: contig of 737 bp in length
            *                    gap of unknown length
            *     1537     2265: contig of 729 bp in length
            *                    gap of unknown length
            *     2266     2972: contig of 707 bp in length
            *                    gap of unknown length
            *     2973     3725: contig of 753 bp in length
            *                    gap of unknown length
            *     3726     4475: contig of 750 bp in length
            *                    gap of unknown length
            *     4476     5247: contig of 772 bp in length
            *                    gap of unknown length
            *     5248     5992: contig of 745 bp in length
            *                    gap of unknown length
            *     5993     6730: contig of 738 bp in length
            *                    gap of unknown length
            *     6731     7474. contig of 744 bp in length
            *                    gap of unknown length
            *     7475     8210: contig of 736 bp in length
            *                    gap of unknown length
            *     8211     8928: contig of 718 bp in length
            *                    gap of unknown length
            *     8929     9640. contig of 712 bp in length
            *                    gap of unknown length
            *     9641    10387: contig of 747 bp in length
            *                    gap of unknown length
            *    10388    11120: contig of 733 bp in length
            *                    gap of unknown length
            *    11121    11879. contig of 759 bp in length
            *                    gap of unknown length
            *    11880    12648. contig of 769 bp in length
            *                    gap of unknown length
            *    12649    13441: contig of 793 bp in length
            *                    gap of unknown length
            *    13442    14200: contig of 759 bp in length
            *                    gap of unknown length
            *    14201    14936: contig of 736 bp in length
            *                    gap of unknown length
            *    14937    15720. contig of 784 bp in length
            *                    gap of unknown length
            *    15721    16404. contig of 684 bp in length
            *                    gap of unknown length
            *    16405    17121. contig of 717 bp in length
            *                    gap of unknown length
            *    17122    17847  contig of 726 bp in length
            *                    gap of unknown length
            *    17848    18601  contig of 754 bp in length
            *                    gap of unknown length
            *    18602    19316  contig of 715 bp in length
            *                    gap of unknown length
            *    19317    20058. contig of 742 bp in length
            *                    gap of unknown length
            *    20059    20787: contig of 729 bp in length
            *                    gap of unknown length
            *    20788    21518. contig of 731 bp in length
            *                    gap of unknown length
            *    21519    22254. contig of 736 bp in length
            *                    gap of unknown length
            *    22255    23015. contig of 761 bp in length
            *                    gap of unknown length
            *    23016    23737: contig of 722 bp in length
            *                    gap of unknown length
            *    23738    24453: contig of 716 bp in length
            *                    gap of unknown length
            *    24454    25182  contig of 729 bp in length
            *                    gap of unknown length
            *    25183    25937: contig of 755 bp in length
            *                    gap of unknown length
            *    25938    26683  contig of 746 bp in length
            *                    gap of unknown length
            *    26684    27431: contig of 748 bp in length
            *                    gap of unknown length
            *    27432    28179: contig of 748 bp in length
            *                    gap of unknown length
            *    28180    28941: contig of 762 bp in length
            *                    gap of unknown length
            *    28942    29667: contig of 726 bp in length
            *                    gap of unknown length
            *    29668    30386. contig of 719 bp in length
            *                    gap of unknown length
            *    30387    31105: contig of 719 bp in length
            *                    gap of unknown length
```

```
*    31106    31889: contig of 784 bp in length
*                    gap of unknown length
*    31890    32643: contig of 754 bp in length
*                    gap of unknown length
*    32644    33422: contig of 779 bp in length
*                    gap of unknown length
*    33423    34169: contig of 747 bp in length
*                    gap of unknown length
*    34170    34901: contig of 732 bp in length
*                    gap of unknown length
*    34902    35643: contig of 742 bp in length
*                    gap of unknown length
*    35644    36403: contig of 760 bp in length
*                    gap of unknown length
*    36404    37119: contig of 716 bp in length
*                    gap of unknown length
*    37120    37831: contig of 712 bp in length
*                    gap of unknown length
*    37832    38555: contig of 724 bp in length
*                    gap of unknown length
*    38556    39332: contig of 777 bp in length
*                    gap of unknown length
*    39333    40130: contig of 798 bp in length
*                    gap of unknown length
*    40131    40891: contig of 761 bp in length
*                    gap of unknown length
*    40892    41629: contig of 738 bp in length
*                    gap of unknown length
*    41630    42372: contig of 743 bp in length
*                    gap of unknown length
*    42373    43156: contig of 784 bp in length
*                    gap of unknown length
*    43157    43889: contig of 733 bp in length
*                    gap of unknown length
*    43890    44610: contig of 721 bp in length
*                    gap of unknown length
*    44611    45411: contig of 801 bp in length
*                    gap of unknown length
*    45412    46177: contig of 766 bp in length
*                    gap of unknown length
*    46178    46940: contig of 763 bp in length
*                    gap of unknown length
*    46941    47711: contig of 771 bp in length
*                    gap of unknown length
*    47712    48462: contig of 751 bp in length
*                    gap of unknown length
*    48463    49258: contig of 796 bp in length
*                    gap of unknown length
*    49259    50004: contig of 746 bp in length
*                    gap of unknown length
*    50005    50781: contig of 777 bp in length
*                    gap of unknown length
*    50782    51487: contig of 706 bp in length
*                    gap of unknown length
*    51488    52199: contig of 712 bp in length
*                    gap of unknown length
*    52200    52928: contig of 729 bp in length
*                    gap of unknown length
*    52929    53736: contig of 808 bp in length
*                    gap of unknown length
*    53737    54547: contig of 811 bp in length
*                    gap of unknown length
*    54548    55305: contig of 758 bp in length
*                    gap of unknown length
*    55306    56070: contig of 765 bp in length
*                    gap of unknown length
*    56071    56822: contig of 752 bp in length
*                    gap of unknown length
*    56823    57607: contig of 785 bp in length
*                    gap of unknown length
*    57608    58386: contig of 779 bp in length.

Query Match          18.9%;  Score 562;  DB 42;  Length 58386;
```

```
  Best Local Similarity   94.5%;  Prod. No. 4.4e-??;
  Matches  637;  Conservative    ??;  Mismatches   31;  Indels    ?;  Gaps    5;

Qy     930 cctagtcactctctgtcccagtattcatttttatttttcaaaatatcactaattcaaattgt 989
           |||  ||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |||||||
Db   56148 CCTCATCACTCTCTGTTTTA-TTTTCATTTTATTTTCTACTTTTT-TTAATTTAAATTTT 56205

Qy     990 attatttaactatttgtttgcatctttattgttttctcttactatcatcaaatttatctttc 1049
           |||||||||||||||||  ||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56206 ATTATTTAACTATTTGTGGGCATCTTTATTGTTTTCTCTTACTATCTTAAATTTATCTTTC 56265

Qy    1050 cacaggagcaaggactttatatttgatcctgtgtctcccattaatgtgggaaaaaaaaaag 1109
           ||||||||||||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56266 CACAGGAGCAACGACTTTATATTTGATCCTGTTTCTCCCATTAATGTTGGAAAAAAAAAAG 56325

Qy    1110 ggatggtttaggggggagtgagagaaatgataaaaaaggaggatatttaaaatttttaaaat 1169
           |||||||||| |||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56326 GGATGGTTTAGGGGGAGTGAGAGAAATGATAAAAAAGGAGGATATTTAAAATTTTTAAAT 56385

Qy    1170 catccaagactgacactcctgatatgaaacttcagtctttctttagttttttttttttcatc 1229
           |||||||||||||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56386 CATCCAAGACTGACACTCCTGATATGAAACTTCAGTCTTTCTTTAGTTTTTTTTTTTCATC 56445

Qy    1230 tatgaaaatctggtttaaggaaattgtgcccaaatgcaattgtttttttttttttttttttt 1289
           ||  |||||||||||  |||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56446 TATGAAATCTGGTTTAAGAAATATGTATAAATACAATGATTTTTTTTTAAATTTATATC 56505

Qy    1290 gtagtttctgtttcagttctagccaaaatggaatttctctaaaatataaatatttaaaagt 1349
           |||||||||||||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56506 GTAGTTTCTGTTTCAGTTCTAGTTAGTTTGAATTTTTTGAAATATATATTAAAAGT 56565

Qy    1350 gaaagcgagagagtctgggacacagctattgttttattcatttttaaaatttatttaaat 1409
           |||||||||||||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56566 GAAAGCGAGAGAGTCTGGGACATAGCTATTGTTTTAATTCATTTTAAAATTTATTTAAT 56625

Qy    1410 cttaaaagcaatacagtgccacaattaagtgatggcctaaattatgaaaaatatttaaatt 1469
           |||||||||||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56626 CTTAAAACCAATACAGTGCCACAATTAAGTGATGGCCTAAATTATGAAAATATTTAAATT 56685

Qy    1470 ctagcagttccactcctagaaagccataatt-agaaaattcaatcaaattaattaattaatt 1528
           |||||||||||||||||  |||||  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56686 CTAGCAGTTCCACTCCTAGAAAAACATAA-TAAGAAAATTCACTAAATTAATTAATTAATT 56745

Qy    1529 tcaaccaatggtcattgag-ccctaaataatgattg aatt t  ata t tt  ttaaatg 1584
           |||||||||||||||||| |||||  |  |  |  |  |  |  |  |  |  |  |  |||||
Db   56746 TTAACCAATGGTCATTGAGCCCTAATTAATCATTGTAAGTTTATAATTTTTATTTTGATA 56805

Qy    1585 gagcttaaggcttt 1598
           |||  |  |  |  ||
Db   56806 GAACTTTAAGGCTT 56819


RESULT   3
AC012249/c
LOCUS       AC012249    58386 bp    DNA                 HTG       21-OCT-1999
DEFINITION  Homo sapiens clone 16_N_23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC012249
VERSION     AC012249.1  GI:6091795
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone 16_N_23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,P., Borowsky,L., Bonkhalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donohoe,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,I.,
```

Galagan,J., Gardyna,S., Grant,C., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Liou,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,P., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Ye,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
*        1      799: contig of 799 bp in length
*                   gap of unknown length
*      800     1536: contig of 737 bp in length
*                   gap of unknown length
*     1537     2265: contig of 729 bp in length
*                   gap of unknown length
*     2266     2972: contig of 707 bp in length
*                   gap of unknown length
*     2973     3725: contig of 753 bp in length
*                   gap of unknown length
*     3726     4475: contig of 750 bp in length
*                   gap of unknown length
*     4476     5247: contig of 772 bp in length
*                   gap of unknown length
*     5248     5992: contig of 745 bp in length
*                   gap of unknown length
*     5993     6730: contig of 738 bp in length
*                   gap of unknown length
*     6731     7474: contig of 744 bp in length
*                   gap of unknown length
*     7475     8210: contig of 736 bp in length
*                   gap of unknown length
*     8211     8928: contig of 718 bp in length
*                   gap of unknown length
*     8929     9640: contig of 712 bp in length
*                   gap of unknown length
*     9641    10387: contig of 747 bp in length
*                   gap of unknown length
*    10388    11120: contig of 733 bp in length
*                   gap of unknown length
*    11121    11879: contig of 759 bp in length
*                   gap of unknown length
*    11880    12648: contig of 769 bp in length
*                   gap of unknown length
*    12649    13441: contig of 793 bp in length
*                   gap of unknown length
*    13442    14200: contig of 759 bp in length
*                   gap of unknown length
*    14201    14936: contig of 736 bp in length
*                   gap of unknown length
*    14937    15720: contig of 784 bp in length
*                   gap of unknown length
*    15721    16404: contig of 684 bp in length
*                   gap of unknown length
*    16405    17121: contig of 717 bp in length
*                   gap of unknown length
*    17122    17847: contig of 726 bp in length
*                   gap of unknown length
*    17848    18601: contig of 754 bp in length
*                   gap of unknown length
*    18602    19316: contig of 715 bp in length
*                   gap of unknown length
*    19317    20058: contig of 742 bp in length
*                   gap of unknown length
*    20059    20787: contig of 729 bp in length
*                   gap of unknown length
*    20788    21518: contig of 731 bp in length
*                   gap of unknown length
*    21519    22254: contig of 736 bp in length
*                   gap of unknown length
*    22255    23015: contig of 761 bp in length
*                   gap of unknown length
*    23016    23737: contig of 722 bp in length
*                   gap of unknown length
*    23738    24453: contig of 716 bp in length
*                   gap of unknown length
*    24454    25182: contig of 729 bp in length
*                   gap of unknown length
*    25183    25937: contig of 755 bp in length
*                   gap of unknown length
*    25938    26683: contig of 746 bp in length
*                   gap of unknown length
*    26684    27431: contig of 748 bp in length
*                   gap of unknown length
*    27432    28179: contig of 748 bp in length
*                   gap of unknown length
*    28180    28941: contig of 762 bp in length
*                   gap of unknown length
*    28942    29667: contig of 726 bp in length
*                   gap of unknown length
*    29668    30386: contig of 719 bp in length
*                   gap of unknown length
*    30387    31105: contig of 719 bp in length
*                   gap of unknown length
*    31106    31889: contig of 784 bp in length
*                   gap of unknown length
*    31890    32643: contig of 754 bp in length
*                   gap of unknown length
*    32644    33422: contig of 779 bp in length
*                   gap of unknown length
*    33423    34169: contig of 747 bp in length
*                   gap of unknown length
*    34170    34901: contig of 732 bp in length
*                   gap of unknown length
*    34902    35643: contig of 742 bp in length
*                   gap of unknown length
*    35644    36403: contig of 760 bp in length
*                   gap of unknown length
*    36404    37119: contig of 716 bp in length
*                   gap of unknown length
*    37120    37831: contig of 712 bp in length
*                   gap of unknown length
*    37832    38555: contig of 724 bp in length
*                   gap of unknown length
*    38556    39332: contig of 777 bp in length
*                   gap of unknown length
*    39333    40130: contig of 798 bp in length
*                   gap of unknown length
*    40131    40891: contig of 761 bp in length
*                   gap of unknown length
*    40892    41629: contig of 738 bp in length
*                   gap of unknown length
*    41630    42372: contig of 743 bp in length
*                   gap of unknown length
*    42373    43156: contig of 784 bp in length
*                   gap of unknown length
*    43157    43889: contig of 733 bp in length
*                   gap of unknown length
*    43890    44610: contig of 721 bp in length
*                   gap of unknown length
```

```
            *    44611    45411: contig of 801 bp in length
            *           gap of unknown length
            *    45412    46177: contig of 766 bp in length
            *           gap of unknown length
            *    46178    46940: contig of 763 bp in length
            *           gap of unknown length
            *    46941    47711: contig of 771 bp in length
            *           gap of unknown length
            *    47712    48462: contig of 751 bp in length
            *           gap of unknown length
            *    48463    49259: contig of 796 bp in length
            *           gap of unknown length
            *    49259    50004: contig of 746 bp in length
            *           gap of unknown length
            *    50005    50781: contig of 777 bp in length
            *           gap of unknown length
            *    50782    51487: contig of 706 bp in length
            *           gap of unknown length
            *    51488    52199: contig of 712 bp in length
            *           gap of unknown length
            *    52200    52929: contig of 729 bp in length
            *           gap of unknown length
            *    52929    53736: contig of 808 bp in length
            *           gap of unknown length
            *    53737    54547: contig of 811 bp in length
            *           gap of unknown length
            *    54549    55305: contig of 758 bp in length
            *           gap of unknown length
            *    55306    56070: contig of 765 bp in length
            *           gap of unknown length
            *    56071    56822: contig of 752 bp in length
            *           gap of unknown length
            *    56823    57607: contig of 785 bp in length
            *           gap of unknown length
            *    57608    58386: contig of 779 bp in length.

  Query Match             14.4%;  Score 406.2;  DB 42;  Length 58386;
  Best Local Similarity   89.0%;  Pred. No. 1.9e-56;
  Matches  517;  Conservative    0;  Mismatches   58;  Indels    6;  Gaps    5;

Qy    1413 aatacagtgccacaattaagtggtggcctgacttttag-agacatgctaatttctagcagt 1471
           || |  | || || |||| ||||||||    ||| || | |||   || || ||  | |
Db   33307 AACAGTGCCCCCCATTTAAATGGTGGCCGACTTTAAGAAAACACCTTATTTTTAATATT 33248

Qy    1472 tccactcctagaaagagcatgattagaaaattcacaaattagctaattatttcaaccaat 1531
           ||| |||||| |||||||| |||| ||| || || | |||  || ||||||||||||||
Db   33247 TCCTCTCCTAAAAGAGCATTATTAAAAATTTCCCCCATTTAATTATTTATTTCAACCAAT 33188

Qy    1532 ggtcattgagccgacatgagcagtgcagttcagactctaccttgacagagcttaaggctt 1591
           |||| |    |  ||| | ||||||||||||||| ||||||||||||||| | |||||||
Db   33187 GGTCTTGACCCGACTACAAGCGTCAGTTCTACCTTGACAGAGTTAAAGGCTT 33126

Qy    1592 tatttcctcaaaaaggaaatggctaaacgacttgttttatagtctattatgcttaagtag 1651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33127 TATTTCCTCAAAAAGGAAATGGCTAAACGACTTGTTTTATAGTCTATTATGCTTAAGTAG 33068

Qy    1652 agttttttttttttccatatagttattgggctacaggtggtatttggttacatgagtaagt 1711
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33067 AGTTTTTTTTTTTTCCATAAGTTATTGGGCTACAGGTGGTATTTGGTTACATGAGTAAGT 33008

Qy    1712 tctttagtgtaagtagagattttatattgcatttttatcatgtgtgaatttctgtagtta 1771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33007 TCTTTAGTGTAAGTAGAGATTTTATATTGCATTTTTATCATGTGTGAATTTCTGTAGTTA 32948

Qy    1772 tctgggtagttgattctgtaggttgaaatacacaaataggaggtgaaagaag--gtctgg 1829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db   32947 TCTGGGTAGTTGATTCTGTAGGTTGAAATACACAAATAGGAGGTGAAAGAAGCCTCCGCC 32888

Qy    1830 aatttgtactatttcctgtcatcccagagagtcttcagcagctcctctatccatcccagt 1889
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   32887 CATTTGTACTATTTCCTGTCATCCCAAAGAGTCTTCAGCAGTTCCTCTATCCATCCCAGT 32828

Qy    1890 tgccccctaccagctcctgctcctttcccctgttaatatattgaccaaatcccacagaaga 1949
```

```
           || ||||||||||||||||||||                                     |
Db   32828 TGTCCCCTACCAGCTCCTGCTCCTTTCCCCTGTTAATATATTGACCAAATCCCACAGAAGA 32771

Qy    1950 ctgtggcatgaccccaactggccccctcttttgttctgtgatat 1992
           || |||||||||| ||||                          |||
Db   32770 CTCTGGCATGACCCAACTGGCCCCCTCTTTTGTTCTGTGTTTT 32728


RESULT   4
AB002365
LOCUS       AB002365      5654 bp    mRNA            PRI       11-FEB-1999
DEFINITION  Human mRNA for KIAA0367 gene, partial cds.
ACCESSION   AB002365
VERSION     AB002365.1  GI:2224675
KEYWORDS    KIAA0367.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
            plus clone:HHC141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Yana, Kisarazu 292-0812
            Yana, Kisarazu, Chiba 292, Japan (E-mail: kaz@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3933,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1. .5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HHC141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .2151
                     /gene="KIAA0367"
     CDS             <1. .2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /product="KIAA0367.1"
                     /db_xref="GI:2224675"
                     /translation="ALPQSDPHLSLENPALVPLALLASLTTLLISFAAPLDHSFSDASG
                     LNTSTGTIDDMSKLLSLSHHETPVDSCLRLLTSASASWNLTSTTVMSQNIDEDLL
                     REPEHFLPGHLPPLAELSLKLLAYLPPLSPLLRLLPALPMPLPPPHSPEDESLGC
                     RAAEIVLSALPDRRREGNLAETKNFLPGSLAYLDITELPDVYTLPVIAGSNILSFSN
                     VDWEVEIDNSLLPASSISPPNGASKEIPLETEKTIPTKDTL_IKSDYKEEPCTEKN
                     EDRHALHMDYILVNREENSHSKPETCNDRDSPLNLLDLVVDAKPTLLSIGLASFPDTC
                     LPASLNESKILVAEINLSALTLPSSFLSLTLLLNNTLTP  AS  SLAFISGPGWS
                     GKIVEPPSELGLSESPILLILEEMKPLESLALEASIPVS_SLKSKSPSRAGPDAVTH
                     DSEWEMLSPQFVCKNMISGTDMEEDTDFLCLGIPIRSPDNCTLATLVDMSFPFDEGVLS
                     PSAADMRPEPPENSLLENSTHFRRIKLTAFNIMLSLELIERNTLDFTLNLESPDEIDINV
                     DELDTFTLAAS.L.LSLL.TLNFDLS.EALSFLEPTDMFPRALNSLNSLNSFVVISEQSQR
                     IDMKVIEPYRRVISHSQTVSLNALIVFAASFLPLSSFAIVHVVMENLFLYVISILE
                     LMVAEDYMIVVLNCATTKKEDARASLEEDMLFLV"
BASE COUNT     1664 a   1188 c   1198 g   1604 t
ORIGIN


  Query Match              9.5%;  Score 282.4;  DB 9;  Length 5654;
  Best Local Similarity   99.6%;  Pred. No. 2.1e-34;
  Matches  283;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      123 gactactgtcagaggatgtaggaatggacatcccctttgaagagggcgtgctgagtccca 182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1469 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCA 1528

Qy      183 gtgctgcagacatgaggcctgaagctcctaattctgtggatcttaatgacactcatcctc 242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1529 GTGCTGCAGACATGAGGCCTGAAGCTCCTAATTCTGTGGATCTTAATGACACTCATCCTC 1588

Qy      243 ggagaatcaagctcacagcccaaatatcaatctttctctggaccaaagtgaaggatcta 302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1589 GGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 1648

Qy      303 ttctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttg 362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1649 TTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTG 1708

Qy      363 ataccccggatgaagcagattcttttgagtactctggccatggt 406
            |||||||||||||||||||||||||||||||||||||| |
Db     1709 ATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGGCCATGAT 1752


RESULT   5
AC009556/c
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15, clone RP11-57C21, map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liew,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pallara,V., Riley,P., Roberts,D., Roy,A., Savery,F.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yu,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165034 bases at least Q20
                Insert size: 165000: agarose-fp
                Insert size: 169709: sum-of-contigs
                Quality coverage: 4.2 in Q20 bases: agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *                   gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *                   gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *                   gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *                   gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *                   gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *                   gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *                   gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *                   gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *                   gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                   gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                   gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                   gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
```

```
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


  Query Match             3.6%;  Score 106.8;  DB 44;  Length 169798;
  Best Local Similarity   90.5%;  Pred. No. 4.8e-08;
  Matches  114;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      18 ggcgaatctgtatttccagttaactgctcagaagagagatgctgaagagctgtagtcgtg 77
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   97871 GGCGAATCTGTATTTCCAGTTAACTGTTCAGAAGAGAGATGTTGAAGAGCTGTAGTCGTG 97912

Qy      78 catccttctcaccctccgttagaaagcctcctctcatcctcagaagactactgtcagagg 137
           |||||||||||||||||||||||||||| | ||||||||||||||    ||      |
Db   97911 CATCCTTCTCACCCTCCGTTAGAAAGCCTCCTTCTCATCCTCAGAAGGTAGGTGGGTCCTG 97852

Qy     138 atgtag 143
           ||||||
Db   97851 GTGTAG 97846


RESULT   6
AC011488/c
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG, HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota, Metazoa, Chordata, Craniata; Vertebrata; Mammalia;
            Eutheria, Primates, Catarrhini, Hominidae, Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *              gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *              gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *              gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *              gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *              gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *              gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *              gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *              gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *              gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *              gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *              gap of unknown length
            *     8340     9880: contig of 541 bp in length
            *              gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *              gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *              gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *              gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *              gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *              gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *              gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *              gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *              gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *              gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *              gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *              gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *              gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *              gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *              gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *              gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *              gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *              gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *              gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *              gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *              gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *              gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *              gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *              gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *              gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *              gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *              gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *              gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *              gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *              gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *              gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *              gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *              gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *              gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *              gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *              gap of unknown length
```

```
            *    41243    41839: contig of 597 bp in length
            *                    gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *                    gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *                    gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *                    gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *                    gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *                    gap of unknown length
            *   112601   174301: contig of 66701 bp in length
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42598 t     85 others
ORIGIN


  Query Match             3.4%;  Score 100.6;  DB 41;  Length 179301;
  Best Local Similarity   64.3%;  Pred. No. 4.1e-07;
  Matches  151;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy      201 ctgaacctcctaatttctctggatcttaatgaactcatctctggagaaatcaagctcacag 260
            | | ||||| || | ||  || | || |   | |||| |  ||| | | |||    |
Db   124306 CAGCTCCTCCCAACAGCTAAATTTCAACCGACCGCATCGTAAGAGGAAGACGCTGGTGG 124247

Qy      261 ccccaaaatatcaatctttctctggaccaaagtgaaggatctattctctctgatgataact 320
            ||||| | |||||||||| || || | || || | || || ||||    ||
Db   124246 CCCCAGACATCAACATTTCTTCTGGATCAGACTCAGGGTTCCTGTCTTCCGATGACTTCT 124187

Qy      321 tggacagttccagtgaaattgacatcaatgtggatgaacttgatacccccgatgaagcag 380
            |||| | || |||||  | || || || ||| ||||  | || ||||||||||| | |
Db   124186 TGGATACCCCTGATGACTTGGATATTAACTTCATGACATCCAGACCCCCGATGAGACCG 124127

Qy      381 attcttttgagtgacactggccatggtaagtcaataagttggcaaggccaaaagtta 435
            | ||  | |||| |   ||  |||| ||   | |   ||| |||| ||  | |
Db   124126 ACTCGTTGAGTTCCTGGGAATGGCAACCAACTGGATTGGAAGGTAAAGTTCA 124072


RESULT   7
HSJ730H16/c
LOCUS       HSJ730H16  158282 bp    DNA             HTG       03-DEC-1999
DEFINITION  Homo sapiens chromosome 1 clone RP4-730H16 map p13.1-13.3, ***
            SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AL122006
VERSION     AL122006.2  GI:6522752
KEYWORDS    HTG, HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158282)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 3, 1999 this sequence version replaced gi:6066846
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.  Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00233  Length: 8500bp
            Contig_ID: 00275  Length: 1268bp
            Contig_ID: 00299  Length: 8761bp
            Contig_ID: 00628  Length: 1023bp
            Contig_ID: 00870  Length: 18527bp
            Contig_ID: 01656  Length: 88357bp
            Contig_ID: 01852  Length: 28201bp.
            * NOTE  This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..158282
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p13.1-13.3"
                     /clone="RP4-730H16"
                     /clone_lib="RPCI-4"
BASE COUNT    48439 a  28295 c  38067 g  40680 t     4801 others
ORIGIN


  Query Match             3.4%;  Score 100.4;  DB 33;  Length 158282;
  Best Local Similarity   52.1%;  Pred. No. 4.6e-07;
  Matches  279;  Conservative    0;  Mismatches  246;  Indels   11;  Gaps    2;

Qy     2213 agaattgtaaactaagttgaatttttaaatctacctaaattttcaaatcaaagatatcaa 2272
            | | || ||   ||  || | ||  |  ||| | | |  |  |  |  |  |  |
Db   151712 ATATTTATATGATATATTTATATTTATTATATTTATTATATATTATATATATTATATATT 151653

Qy     2273 gtcagaggttgaaaactttgaaaaacaaaatgcctttgaatagctatttttcaaaaat 2332
            |      |  |  | | | |  | |   |  | |  ||  |||||||  | | |
Db   151652 ATATTTATTATATTATATTATATTATATTATATATTTATATTTATATATTTATAATATAT 151593

Qy     2333 gaggacatggataagtgactgtataaaaaaattggattataaatatattttttttttaa 2392
            ||  |||| |  | |||      |||  ||  |  |  |  |  |  |  |  |
Db   151592 ------ATATATAAATATATATTTATAAGATATATTTATAATATATATATATATTAT 151539

Qy     2393 gtatgaaataaatgcttgaaagtaggcttttttttaggttttttaaaaaaaaaaaaaaaa 2452
            |||  |||| ||  | |                 | ||  |  |  |  |  |  |
Db   151538 ATATAAAATATATAATATATTATTAATATAATATATTTATTATATATTAATATATATATAT 151479

Qy     2453 ccaggaatgtacattctaaacaaatttgggttttttaaggttatagttttttaaaaaaaa 2512
            ||  ||| || ||| |||| | ||||    ||| |||   |  |  |  |  |  |
Db   151478 TCAAAAATATATATTATAAATATAATATATATTTATAATATATATTATATAATATATATT 151419

Qy     2513 tgggccataatttatttttcttttaattaaatacttttttcatttatattatatttttgaat 2572
            |    |||  |||| | | | |  |  | | |  |  |  |  |  |  |  |  |
Db   151418 TATAATATATATTATATAATATATATTTATAATATATAATATATAATATATTTATAATAT 151359

Qy     2573 atatatttttcat-----tatatatgttttcattatatattatatatttgtatttgtatttt 2627
            ||||  | | ||     | ||||  |  |||||||| |||||||| || |||  |  |
Db   151358 ATATTATATAATATATATTTATAATATATATTATATAATATATATTTATAATATATATAATA 151299

Qy     2628 ttcatttatatagttttcattgtatattatatacctcattatattatatatttattatta 2687
            |  || |||||  | ||     |||| |          ||  |  |  |  |  |  |
Db   151298 TAAATATATATTATATTATAAATATAATAATATATATATATAATATAATATATTTATATATA 151239

Qy     2688 tagttttcatttaatttatttcatttttataactagaacaatgaaagcacaagaaaaat 2747
            |  | | | || || | |     ||||  |  |  |  |  |  |  |  |  |
Db   151238 AATATATTATATATTATTACATATTTTATTATATATATATTATTACATATATAAAAT 151183


RESULT   8
PFAVAR23A/c
LOCUS       PFAVAR23A   22243 bp    DNA             INV       14-SEP-1995
DEFINITION  Plasmodium falciparum (strain FCR3) variant-specific surface
            protein (var-2, var-3) genes, complete cds's.
ACCESSION   L40609
VERSION     L40609.1  GI:886376
KEYWORDS    variant-specific surface protein.
SOURCE      Plasmodium falciparum (strain FCR3) (clone PFVAR23) blood stage
            DNA.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 22243)
```

```
  AUTHORS   Su,X.Z., Heatwole,V.M., Wertheimer,S.P., Guinet,F., Herrfeldt,J.A.,
            Peterson,D.S., Ravetch,J.A. and Wellems,T.E.
  TITLE     The large diverse gene family var encodes proteins involved in
            cytoadherence and antigenic variation of Plasmodium
            falciparum-infected erythrocytes
  JOURNAL   Cell 82 (1), 89-100 (1995)
  MEDLINE   95330813
FEATURES             Location/Qualifiers
     source          1..22243
                     /organism="Plasmodium falciparum"
                     /strain="FCR3"
                     /db_xref="taxon:5833"
                     /clone="PfYED9"
                     /dev_stage="blood stage"
     gene            427..10322
                     /gene="var-3"
     exon            <427..9217
                     /gene="var-3"
                     /number=1
     CDS             join(427..9217,9093..10322)
                     /gene="var-3"
                     /note="specific for blood stage; putative"
                     /codon_start=1
                     /product="variant-specific surface protein"
                     /protein_id="AAA75292.1"
                     /db_xref="GI:886277"
                     /translation="MAPAQGGTNKTAKEVLDEIGKEVQEKATEDALTYPNDLQGNLSQ
                     AKFHGVPIDVKNPCDLNYEIHTNVVKGPKKENPCPGPEEKPFSDVLSGQCAKNTIKDS
                     VTNSVGACAPFPPLHVCDPNLELIKTDENTSTHDILVDVLLTAKHEGESLVKHYKEYI
                     KFNPNFNIGTVLAPSFADIGDIVPGKDLYLIYDDEEKEKPKQLEKNLKKFFQKIHDDV
                     MKTSGPTNGKKSAEAQKPYNDATGNYYKLPEDWWNANPDQVWKAITGDAADNDEYFEN
                     SSDGLYVFSNGQCGPNEGKVPTNLDYVPQHLPWFDEWAEDFCPKPNITLKSAKEKCGG
                     EGNIKYGSLNGYDGTKVFEKKDSGSSDGNCTACSNICIAYDAWLPNQPNEFEKQKIKY
                     IKEIEKYKSSSDKSNSNIGNKYYNEFYENFSKKEYETLQNFLKLLNKSMVCQEKIEEE
                     EVIDFNKDEDMVFHPSEYCQPCPDCVVQCKGGKCTEDKKNDKCPSKIIKFILQSEEPT
                     EIHVLNSDDKQGDITKKLEVFCSSIINYEGPNVQKWKCYNKNSDYNNCEMNISSYKDS
                     TDANVMLSVECFHSWAKNLLIDTIKWEHQLKNCINNTNVTYCESKCIKNCECVEKWIK
                     FFEHEWEKVKNVFGNNNPMSYIYYNNLSPVFDSFLFQVMFALDQDEKGKWDQFTEDLY
                     KKFEPSKTNTPTGKSQDAIFLLDHLKDNALTGPDNNSNESQDVSKKVKTNPQGKNPS
                     ASNNLVRVKFLAEMMQFYARFQLEKPGGEINLKGDATKGTYPQGGPADGFKNVQSINQ
                     NHINVQNNNPAYTYQGPGTPGVTGSNGGVPMKITGTPWKPGPQIQMSAEDTYMPPPPQHM
                     CISNLEYLQTKDGPLKQGPGPNLVNNSFLGDVLLSAKMDAGKIIELYKKQNNKSNLTDP
                     EDNESACPALPYSFADLGDIIPGPDLWDKNSDAKPLQTNLKEIFTKIKEELPEDIKKK
                     YDKDGTDHKLLPEDWWEANPHQVWPAMKCAIENDKDMKCNGIPIEDYIPQRLPWMTEW
                     AEWFCKEQSRLYNKLVADCKSCKGKAKSCTQKDGDCTKCKAACDNYNKKIKPWEEQWE
                     KIKNKYAQLYKKALDSVNGKEESKKKTASDAKDQQVVHFLAELIPKSGGGKGGKNVKT
                     TVSPITTPNTLYSSAAGYIHHELGPTVGCNTQKEFCYSKNGKYAFKDPPKGYEEACKC
                     NDRNPFPQPAPKKEDEDACDVVKPLLKDKGETDDIDGCNQKYKAGKDKYPGWDCNSQI
                     HTTHNGACMPPPRQKLCVSGTTKTDPIKAIEYIPTEFIKSAAIETHFAWDPYKEDNGE
                     AEAELKNGNIPEGPKPQMYYTPGDYPDIFFGPDISTHAYISGVSPKVTTILEKENDAK
                     YAAKQNSNNELLDDWWDQHGKDIWEGMLCALTHKISDEEKKKEIKNKYSYKKLNESPK
                     GSNKVEDFAKKPQFLPWFIEWGDEFCAQPEEKEAKVKVSCSDAKDYDGCKNTKSNASC
                     VSACKVYEDYITKFKVEYTKQKGKFDAEKITDKFGYEGPSTKDASEYLKKFQLDDTDN
                     CMQKVKNNTEYWNTPNKTYTNSNLEKPCECQPPQEPPPGPEGGAPSDSGPPDTTPPPA
                     GSDAPSNTVPSPPPPAGDTVHEVAEVQEPEEEEDEDGDLPEDQDEDVEVAGAEEEEDL
                     DVGVAPILPGPTNSPDEDEIEASEEEDDDDDDAQDITEVTGQGEEEIAEDHQDIIEEI
                     VDQEKAEEDKDGGGETPQKFIQPKVEVNPCDIVKTLFTTTETLKEACPTKYVNGPEKF
                     PNWKCISSGSDASGSICIPPPPPKLYLHKIEGVDTTVSSDGFTTTPITHDALPEAFTQ
                     TAAVETFFLWHPYKKIKEKEPQEELQNGTFLLPPAQKVSPEDNPEHPQKKLKEGKIPE
                     EFKPQMFYTLGDYPDLGVGVKDDVAQALEASGDNKSGDKNIKDISEKIKSVIEYSGEQ
                     IPPGPKPGQTITKPEEWWQKNQEHIWNAMICALTHNTDTPQVDDQVKGQLPENGKNTF
                     KNSQYQYKNVTISSVSSNGGPIGNIKLEQFASPPTFLPWLEEWGEEFCPKQYHKLEPI
                     KEECHKDGNPNCDDDGPECKPMCPCKDGSPETLKCPSCAKSCKSYKKWISPKKDEPTY
                     QKGAYEKQKKDAEGNNNDYKEFSKTLPNYNDAAAFLNSLKNGPCSKNDDSVQDEIKF
                     DDEPKTFGHETYCKPCSKITVKCKENNHCDNSKPNDCRNINSISAEDIEKPSNSTQDV
                     TMSVSDSNTNGNKFYDLNDCIKAGIFKGIPEDVWKCGEYCGVDICTLEKTNNEEPVSA
                     KENDNKNQIILIPVLFKRWIFSFLEDYNKINDKISHCMKNDKKSPCINGCQNKCNCVE
                     KWIEKKKSEWGKVPEPYINQYPDKNSNEAFEVKSFLETLIPQIPVVTDKGKHDSLIQL
                     KKLLKCNCSEKSENSNEKDVVLCLLKKLEDKAKNCKDQASGEPCPQTTSENPDDEDIL
                     LEEENPVEAPNICPKVEEPEPVVEEEKCDLAEAPSKESSTEENSGEGSNSEQNPPSKP
                     EEEPPPPTSSETDTPPPAPPIIQPSQADQPTNSISDILSSTIPFGIAIALTSIVFLFL
                     KKKTKSSVDLLPVLNIPKGFYGMPTLKSSNPYIPYASDFYKGKTYIYMEGDSDSGHYY
                     EDTTDVTSSESEYEELDINDIYVPGSPKYKTLIEVVLEPSGNNTTASGKNIPSDIQND
                     IQNDCIPSSKITENENNTLKDDEIQQLQELENDVDNEVTGDNGKTNTNITTTSRHN
                     VEEKPTIMSIHERNLITGEEINTNYNVNTMGTTINRENNVTGGTLLTNDALNGDYD
                     IYDENLPFPENELPGTNNTY*TTTGPVKPPAT*TTTQ*LLPTPVLGPLFLMIEKLK
                     NHHEPLAPLPE*NNLIGNIEE*LNATLN*SSGNNNLNAIINETPNMDTILD
                     DLDINNEPYYPVQEDIYDVNDNETSTVLSNDMIYGPVLIEMSVNTKLYKEKYPIRD
                     VWDI"
     intron          8218..9092
                     /gene="var-3"
     exon            9093..10322
                     /gene="var-3"
                     /number=2
     exon            <11749..18839
                     /gene="var-2"
                     /note="putative"
                     /number=1
     gene            11749..20741
                     /gene="var-2"
     CDS             join(11749..18839,18938..20741)
                     /gene="var-2"
                     /note="specific for blood stage; putative"
                     /codon_start=1
                     /product="variant-specific surface protein"
                     /protein_id="AAA75295.1"
                     /db_xref="GI:886278"
                     /translation="MVLQPAVGGGDTLLKVAVLLLVAVLLLSLAALLDALLPSNG
                     KLKSTLSLAIPAAIDDPLTLLFLGLMLMVLSILALLNLALRVRFSYIEG
                     AECDKSKIPGSNSNKDDCACAFPPLHLCDPNLKFLKTTENNLLADVTEAAKFEAE
                     SLEKYPGQVTLVNLGSPNNRTIEELRPPLDSIVSRELLPRNLAVLEELKRKI
                     PRKIVCNLNEAVLPRYPELKDTPNSVRLFNLVLENVLNPLVAAGGIYFRQ
                     TGTTGTWTMEPVSLVNLFLFLVVLLFLVKINGRNY
                     SRNLKDSGNGLLGEETDVETPHVTGEEPKKPFWTIVSLLNLLLK*KKKCE
                     NEMLSKSKFKQPPVNVTDSVKEPVNLRSENVSGLKPSLELLNEEEELQEFSNDLG
                     KIDFKTVDGGPVPAIPGPASLPNNSNVTTGNGTTSNVTPEPSGTVLLPGVNLKRNKGEC
                     KGGKRLNIFRGLNLVILAIGLFALLLLIMSAVSAAIIQWKGTYGDK
                     EYEVGTLENPNKPEEPEEELKPFVNTTVWIVLLNETVNETFNLPNVTEKAKEGKC
                     KNECKTDGGPVPAIPGPASLPNNSNVTTGNGTTSNVTPEPSGTVLLPGVNLKRNIDELFQD
                     ITEAYGNSQKLIERETLAKKTQEALDATLKNTTELLETTLSAVQGKKIQEECQ
                     EKKETPVPNPCVNNTTCALCWTVLKLWTVLLLLLLLLLLLLLLLLLLLLLWKYSNQG
                     DPDDFKKNLCGTTVPWNATLNKSNETNNTWHLFNETINASHNGTDLYLPP
                     RRIHFTTENLEKLNETRPTGLLTPFVLLPIFLLVLLALLAVAFCLIKRRQIKDNSSIC
                     RTMFYSPATTLVPLSLNLSNCLACVDVDNETLPKPTTSWMFLQLKFNCNDKY
                     TNDSKHTKLRSLVDANPDQVTKAITTEHLGYSFLSDGWMVEWAEW
                     FIPEVPEIVFFLLDIFGINLSKNSLYELVFVANLVMVLVSLFLLNELSNKYQ
                     ILIWQAKIAAINSTERSTTTEEVKDKVNQFLRLVDKNITSLTKPSFEAHPDSRPRR
                     AATSKSCVYETTNNYIKQEAKTPETLLNVTQNNNCNNNSNISLIETNFKGKMCNEN
                     KASSPEELGRSLSPESNLTPTEELEVNSSEEDLELDPELELKLNNNNVHDGADEKA
                     GAVSQPEASPTTKDVYYTLCIVKELTGNYCTLFKDISTKIFCKNTCNRCIPTKISN
                     DVTGETSQFRNWSETNSETNFLVADTIIEKLNLSDESERKNHNIETSVS
                     QGEAIEAPQSEATAVSKESPSGGPETPSLKLSNFLSPLNYTTFIMERYKKEWLAQ
                     KPAELQNTSLPVSTTSTPVPQLKVLNSLPLSLMLILVLVHGGNIS
                     DSGNTNGSNNNNIVLDASCNKDMLKIGEKIELSECNGTTLTPDKWWNE
                     HAESIWRGMIGALTTTFTNFLFSALSLENKLAFLPLSALAKSIASIPTGT
                     YKTQYLYFYVKLSTPAKTPASPTTLSPTMSTPTTPTTCKEPFKRL
                     FIIKEECMIGSCKVTGSCSLSPSTCNELKTVCLDTTLNLSNVKIKSRKKEY
                     DPCANAYSKQKTPVEESPSPSTLNCENVPTPVKLSTTSAVLLAVLSFCKKLNEN
                     GSNDINFGNTELLNLALNLALLGATAILVLAVLAVLVIIIGNFNT
                     MSIIIEVVMETSLKNANLTLTSLLLVLLSLFLLTLHIPLEK
                     GNGALNCGLLVPALLRFWTPFLLKKRIKKLKLSLANLVLLCNKKCNRV
                     GEWIKKKEWKVPEPYKNVLLEVMKEPKNVTPLVTNLLETLVLLNLKPEQTSLDK
                     LKTSLSCNIFDMSKPNCSNETCADLSNNMPLCFDNLPLLLNLPHFNCGS
                     NFPPDEEDLLLEEENVPEPPPPTLLEPETELRIRLEPKALRAELPEQPAEEDG
                     GAIVPSGPPGSGPALKGIVKALEIPKGLLPLALVLFLVLLAWSVGIGFAA
                     FTYFPLFKYTKPVNVLTLTPYSNLLYLMSRISGDSVDIGLTAKNLKPQEIILEIE
                     SGIDSGYIDKVSLLLVLSLLLLLMLINLLLILKLAGLLVVLDPCNNTIASGNN
                     TTASGNNTIASGNNTTPSGTTPSGNNTTPSGTTPSGNNTTPSGTTPSGNNTTPSGQSEPNTEP
                     NMLGVNVDNNTIPSTSGNNVTMKFIMGIDEINLSLGEFFVLEKVYLYLRVLWLVVGVCLIGNNPINISDST
                     NGMDSLTSNNRSFNNDFVTLLSIIFIMSAFSSNSVSFLNLLFGTFHHIK
                     HINIYNVAKPAMLSRIINLNLFHKWLFRNFLMLSFKKKNNGFSLKKLWENETHSG
                     DINSGIFSINNVLNTIVPTLIMFNFAFNNFFNMLIFKEKYNEPY
                     YYDPYEDDIYIHEVDVEKSSMDSIFYVDNNYTNNYNYNYNYKKKLDDNIVNNKKEIFEE
                     EYPISDIWNI"
     intron          18840..19837
                     /gene="var-2"
```

```
     exon            19338   >20738
                     /gene="var-2"
                     /note="putative"
                     /number=2
BASE COUNT     9026 a   2935 c   3959 g   4324 t
ORIGIN


  Query Match             2.9%;  Score 85.4;  DB 34;  Length 22243;
  Best Local Similarity   48.6%;  Pred. No. 9.3e-05;
  Matches  237;  Conservative    0;  Mismatches  251;  Indels    0;  Gaps    0;

Qy    2233 attttaaatctacctaatttttcaaatgaaagatatatagtcagaggttgaaacttttga 2292
           ||  | |  || ||  |   | || |  |   | |||| |  |       |  | |  |  |
Db   19317 ATATGTGTATGTATATATGTATTTAGGTATTATATATGTGTTTGTTAATGTATATATGTA 19258

Qy    2293 aaacacaaagtgcctctgaatagctatatttttataaacatgaggacatggataagtgact 2352
           | | |  | |  |||  |||| |  || | ||  | |  || | | |||  |
Db   19257 TGTAGTTATGGGTATATATATATATATATATATAATATATATATGTATGTGTAAATGTGTAT 19198

Qy    2353 gtataagaggattggatttataaatatgtttgagtttctgaagtatgaaataaatgcttgaa 2412
           |||  |||  |||| |||| | ||  |  |  || | || |   | | | |
Db   19197 GTAGTTATGGGTATATATATATATATATATATGTATGTGTAAATGTGTATGTAGTTATGGGT 19138

Qy    2413 agaagcatgtttcttaggtttctcaatgcgac atttgaaccaggaatgtacattctaaa 2472
           |     | | | |  | |    | | |    |  | ||       | |   || | |
Db   19137 ATATATATATATATATGTATGTATATGTGCATTTAATTAAATTTTTTTATTTTTTTTTAT 19078

Qy    2473 caaaatagggtttttaaggtatagttttcaaaaacatttgggccataatttatttttt 2532
           | |||  |||||| | |  ||  | |  | ||||    | |||  ||||| |
Db   19077 TTCATTAAATTTTTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTT 19018

Qy    2533 tttaatgaatagttttcatttatattatatttttcgtgtatatatttcattatatacg 2592
           ||   | | ||  ||||  ||  |  |  |||||  | | || ||     || ||
Db   19017 TTATTTAATTAAATTTTTTTTATTTATTATTTTTTTTATTTAAATAAATTTTTTAATTT 18959

Qy    2593 tttcattatatattatatatttgtactatatatttttcatttatatagttttcattgtat 2652
           ||  |  | |||  |||| | |  |  |  |  ||  | |  |||| ||| |||
Db   18957 TTTTTTATTTTATGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTATTTTAT 18898

Qy    2653 attatataccttcattatattatataattcattatatagttttcatttaatttatttcatt 2712
           ||||||  |  || | || ||  ||  || | |  |  ||  ||  | | |||  |
Db   18897 CATATATATTTTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTA 18838

Qy    2713 ttataact 2720
           |||||||
Db   18837 TAATCATT 18830


RESULT    9
AC006279/c
LOCUS       AC006279   314303 bp    DNA              HTG       23-SEP-1999
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 16 unordered pieces.
ACCESSION   AC006279
VERSION     AC006279.4  GI:5919277
KEYWORDS    HTG; HTGS_PHASE1
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 314303)
  AUTHORS   Hyman,R.W., Fung,E.I., Qin,F., Rowley,D., Tamaki,T., Kurdi,O.B.,
            Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 314303)
  AUTHORS   Hyman,R.W., Fung,E.I., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Sep 23, 1999 this sequence version replaced gi:4558576.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     7841: contig of 7841 bp in length
            *    7842     8041: gap of unknown length
            *    8042    67772: contig of 59731 bp in length
            *   67773    67972: gap of unknown length
            *   67973   142824: contig of 74852 bp in length
            *  142825   143024: gap of unknown length
            *  143025   147294: contig of 4270 bp in length
            *  147295   147494: gap of unknown length
            *  147495   153406: contig of 5912 bp in length
            *  153407   153606: gap of unknown length
            *  153607   161784: contig of 8178 bp in length
            *  161785   161984: gap of unknown length
            *  161985   163895: contig of 1911 bp in length
            *  163896   164095: gap of unknown length
            *  164096   173907: contig of 9812 bp in length
            *  173908   174107: gap of unknown length
            *  174108   177106: contig of 2999 bp in length
            *  177107   177306: gap of unknown length
            *  177307   179419: contig of 2113 bp in length
            *  179420   179619: gap of unknown length
            *  179620   189515: contig of 9896 bp in length
            *  189516   189715: gap of unknown length
            *  189716   208243: contig of 18528 bp in length
            *  208244   208443: gap of unknown length
            *  208444   237110: contig of 28667 bp in length
            *  237111   237310: gap of unknown length
            *  237311   269228: contig of 31918 bp in length
            *  269229   269428: gap of unknown length
            *  269429   310791: contig of 41363 bp in length
            *  310792   310991: gap of unknown length
            *  310992   314303: contig of 3312 bp in length.
FEATURES             Location/Qualifiers
     source          1..314303
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="3D7"
BASE COUNT   124580 a  33599 c  28900 g  121315 t    2307 others
ORIGIN


  Query Match             2.9%;  Score 85.4;  DB 41;  Length 314303;
  Best Local Similarity   47.0%;  Pred. No. 7.5e-05;
  Matches  262;  Conservative    0;  Mismatches  296;  Indels    1;  Gaps    0;

Qy    2187 tgttttaggtcttaaatgcaaaggaaaaaattgtaaaataaggttaaattttgaaatatt 2246
           | | ||   | ||  |   ||  | | | ||  |   |  ||  | |  | | |
Db 181580 TATATTTTTTATTTTAATATAATAATACACTTTTTCTTTATTTTGCTTTTGTAATTTTTA 181521

Qy    2247 cttaatttttcaaaatgaaagatatatagtcagaggttgaaacttttgaaaaccaaaaatatt 2306
           |   ||  |   ||  |  |||     |      ||      ||  | |
Db 181520 TTATGTTGTTTTATTTAGTATAATATGTAGATATTTTTTGTTTTTATATATATATATTTA 181461

Qy    2307 tctgaatagctatattttgtaaaaatgaggaacatggaataaatgaatgtatatgaaaaaatt 2366
           | |||  |  | |||  | | |  |  |  |  |  |||  |  |  |  |  |
Db 181460 TTTGATAAATATGTATATATTTATGAGATATAAATTTCATATATATTTCATAATTTATATTT 181401

Qy    2367 gattataaatatgtttgagttctgaaatatgaaaataaatatcttaaaaccaaaaaaaaaaa 2426
           | |  |  | |||      |  | |  |  |  |  |  |  |  |  |  |  |
Db 181400 TTCTTTTTAATTTTTTTTTAAGTGTTATTTGTTTCTATATTTATGCCTATGCATTTATTTTA 181341

Qy    2427 taggtttctcacgcgaccatttgaaaccaagaaatgtacatttctaaacaaaaaaaaaaaaa 2486
           ||  || |     |   | | ||  | |'  |  |  |  |  |  |  |  |  |  |
Db 181340 TATCATTATATTTTTATATATATTAAGTGTCAAAAAGAAAAAAAAAAAAAAAATTTTTTTA 181281

Qy    2487 ttaaggtatagttttttaaaaaaaattttaaaaatatatttatttttttttttaaaaaaaaaa 2546
```

```
Db 181280 TTAAAATATTTTTTCTTCTGTAAAATGTTTTCTTTTTTTGTTTATATTTATATTTATATAT 181221

Qy   2547 tttcatttatattatatttttcgtgtatatattttcattatatacgtttcattatatatt 2606
          ||| ||| ' || | |||||     ||| | || |    |||  | ||| || |||||||
Db 181220 TTTTTTTTTTTTTTTTTTTTTTTGTTATTTTTATATATATGAAAGTTGTATGATATATT 181161

Qy   2607 atatatttgtactatatatttttcatttatatagttttcattgtatattatatacctcat 2666
            | |||| :  ' | | |||||  |||    | ||||   | |||   || ||  | ||
Db 181160 TAAAATTTTTTTTTTTTTTTTTTTTTTTTGGTAATTTTTATATATATGAAATTTATATGAT 181101

Qy   2667 tatattatataattcattatatagttttcatttaatttatttcattttataactagaaca 2726
            ||  ||  ||  || ' |  |||| ||  ||| |   |||||||   || |
Db 181100 ATATTTCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATATTTTATTTGTATACAC 181041

Qy   2727 atgaagcacagaggaagtta 2745
          ||  ||  | | | || |||
Db 181040 ATTTAGGAAAAAAAAATTA 181022


RESULT  10
AC011718
LOCUS       AC011718    94387 bp    DNA             PRI       13-OCT-1999
DEFINITION  Homo sapiens, complete sequence
ACCESSION   AC011718
VERSION     AC011718.1  GI:6041881
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94387)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 BAC Clone b2280111 In DGCP Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94387)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
FEATURES             Location/Qualifiers
     source          1..94387
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    26431 a  20252 c  20513 g  27191 t
ORIGIN


  Query Match             2.8%;  Score 64;  DB 40,  Length 94387;
  Best Local Similarity   49.7%;  Pred. No. 0.00016;
  Matches  262;  Conservative    0;  Mismatches  270;  Indels    6;  Gaps    1;

Qy   2189 ttttaggtcttaaatgcaaaggacagaattgtaaactaagttgaatttttaaatctacct 2248
          | | |  |  || ||  |   | | ||| . |  ||   |  | | |  |||
Db  49274 TATAATATAATACATAAATTATATATAATTATACATATAATATAATATATAATTATATA 49333

Qy   2249 taattttcaaatgaaagatatatagtcagaggttgaaacttttgaaaacacaaagtgcctc 2308
          ||||| | | ||  || |||     | | |  | || | | | |    | | |
Db  49334 TAATTATTACATATAATATAATATATAATTAGATATAATTATTACATATGATATAATATA 49393

Qy   2309 tgaatagctatatttttgtaaacatgaggacatggatgaagtgactgtataagagggttgga 2368
          | | ||  |||| ||  || | ||||     |  ||        | |||||      |
Db  49394 IAATTATATATAATTATTACATATGATATAATATATATAATTATATATAATTATATAAATT 49453

Qy   2369 ttataaatatgttttgagttctgaagtatgaaataaatgcttgaaagcagcctgtttcttta 2428
           |||| | | | |     | | | |||  ||| ||  |  | | |   | | |  ||
Db  49454 ATATATACAAGATAATAATATATAATATAATAATATATTACATAATATTTTATATAATA 49513

Qy   2429 ggttttctcagcggaccattttgaaccaggaatgtacattctaaacaaaatagggggttttt 2488
            ||  |  |  | | ||    || ||     ||| | |||    | |
Db  49514 TATTATACTATATATATTATATAATATTTTATATA------AAATATTATATATTATATA 49567
```

```
Qy   2489 aaggtatagtttttcaaaaacatttcagcatatatttatttctttgatgaaatatttt 2548
          |  |  | ||||  | ||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db  49568 TAATATAATATTTTATATAAATATTATATATTATATATAATATAATATTTTTATAAATA 49627

Qy   2549 tcatttatattatatttttcgtgtatatattttcattatatacgtttcattatatatt 2608
          |  | | |  |  |||| |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db  49628 TAATATATAATATATATTATATTATCATATGTTATACAATAATATGTTATATTATAATAT 49687

Qy   2609 atatttgtactatatatttttcatttatatagttttcattgtatattatatacctcatta 2668
          |||||  |  |||| || |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db  49688 ATATTATAATCTAATATATTATATATTATTATATATTATATTATATATAACATTATATAG 49747

Qy   2669 tattatataattcattatatagttttcatttaatttatttcattttataactagaaca 2726
          |||||||| || || ||  |  | |  |  |  |  |  |  |  |  |  |  |  |
Db  49748 TATTATATATTTTATAATAATAATATAATATAATATTATATATTATATATAATATA 49805


RESULT  11
AC005077
LOCUS       AC005077   237090 bp    DNA             HTG       12-JUN-1999
DEFINITION  Homo sapiens clone RG228D17, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   AC005077
VERSION     AC005077.1  GI:3212914
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 237090)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237090)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a "working draft" sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    10638: contig of 10638 bp in length
            *   10639    10655: gap of unknown length
            *   10656   237090: contig of 226435 bp in length
FEATURES             Location/Qualifiers
     source          1..237090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG228D17"
BASE COUNT    62344 a  58300 c  57857 g  58589 t     17 others
ORIGIN


  Query Match             2.8%;  Score 64.6;  DB 33;  Length 237,090;
  Best Local Similarity   50.8%;  Pred. No. 0.00018;
  Matches  257;  Conservative    0;  Mismatches  249;  Indels    1;  Gaps    2;

Qy   2225 taagttgaatttttaaatcta.....tttcaaatgaaagatatatagtcagaggttgga 2284
          || || ||| | | |  ||                                  |||||
Db 222127 TATATTTAATATATATAAATATTTTATATTTTTATTTTTATTTTTTTTTTTTTTTTTTT 222186

Qy   2285 aacttttgaaaacacaaagtgcctctaatgaatatttttatgtaaatatgtttgaaaaaa 2344
          |  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 222187 ATATTTATATTTAATTATATTA-TTTTTTTTTTATTTTTATTTTTTTTTTTTTTTTTTT 222245

Qy   2345 aagtgactgtataagagggttgaattataaatatgtttgaaaaaaaaggtatagaaatgaa 2404
          |  | | | |  |  |  |  ||  |  |  |  |  |  |  |  |  |  |  |
```

```
Db 222247 AIATAAAIATTAAIATATATTTCAATATAAA.TTAATGTATATTTTCATATAAATAATA 222306

Qy   2405 tgcttgaaaagcagcctgtttcttaggt·····ttctcacgcgaccatttgaaccagga 2458

Db 222307 IAIIIAAIAIATTIATATATTTTAATTIAIAIATATTIATATATCTATATATAAATATAG 222366

Qy   2459 atgtacattctaaacaaaatagggtttttaaagtatagttttcaaaaacatttgggcc 2518

Db 222367 ATATATATTTTAATATTTATATATTATAAATATTTATATATTTTATATATTTATATAT 222426

Qy   2519 ataatttatttctttttaatgaatactttttgatttatattatatttttcgtgtatatat 2578

Db 222427 TTAATATATAAATATTTATATATTATATTTAATATATAAATATTTATATATTAAATAT 222486

Qy   2579 tttcattatatacgtttcattatatattatatatttgtactatatatttttcatttatat 2638

Db 222487 ATAATATATAAACAT····TTATATATTATATATTTTATATATTTATATTTAATTATGTT 222542

Qy   2639 agttttcattgtatattatatacctcattatattatataattcattatatagttttcatt 2698

Db 222543 ATATATTAATATATATTTAATATATTTATATATTATATATATTATTATATATATTTATAT 222602

Qy   2699 taatttatttcatttttataactagaa 2724

Db 222603 ATATAATATATATTTTGAGACGAGTA 222628


RESULT  12
AC011293
LOCUS       AC011293   159627 bp    DNA             HTG       04-DEC-1999
DEFINITION  Homo sapiens clone RP11-75F5, *** SEQUENCING IN PROGRESS ***, 1
            unordered pieces.
ACCESSION   AC011293
VERSION     AC011293.2  GI:6524394
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159627)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159627)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Dec 4, 1999 this sequence version replaced gi:6012156
             Center project name: NH0075F05.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   159627  contig of 159627 bp in length.
FEATURES             Location/Qualifiers
     source          1..159627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-75F5"
BASE COUNT    [RP11-75F5 a: 23519 c: 19947 g: 19711 t:
ORIGIN


  Query Match             2.9%;  Score 83;  DB 44;  Length 159627;
  Best Local Similarity   49.4%;  Pred. No. 0.0062;
  Matches  280;  Conservative    0;  Mismatches  287;  Indels    6;  Gaps    2;
```

```
Qy   2182 taggtcttaaatgcaaaggacta[illegible]taggttt[illegible] 2251

Db  74346 TATGTTTTATATATATATAAATATATTATATATATTTTTATATATAAATATATATTTT 74405

Qy   2251 ····attttcaaatgaaagatatat[illegible] 2306

Db  74406 ATATATTTTTATATATAATATATTATATATATTTTTATATATAATTATATAAATATATA 74465

Qy   2307 tctgaatagctatattttgtaa[illegible] 2366

Db  74466 TTTAAGTATAAATATATAATAAATATAAATATATTTTTATAT-AATATATTTT····· 74524

Qy   2367 gattataaatatgtttgagttctgaaatatgaaaataaatgcttgaaaatcatgt[illegible] 2426

Db  74525 TATATATAATATATTCTATATAAATATATATTTTATATTATATTTCTATATAATATATATTT 74584

Qy   2427 taggtttctcaggcgaccattt[illegible] 2486

Db  74585 TATATTATATTCTATATAATATATATTTTATATTATATTATATATTATATATTTTATATT 74644

Qy   2487 ttaaggtatagtttttcaaaaacatttgggccatagtttatttctttttaatgaatact[illegible] 2546

Db  74645 ATATTATATAAGATATATTTTATATTATATAATATAATATATATTTAATATTATATTATA 74704

Qy   2547 tttcatttatattatatttttcgtgtatatatttttcatttatatacgtttcatttatt[illegible] 2606

Db  74705 TAAGATATATTTTATATTATATTATATATAATATATATTTTATATTATGTTATATTTAAT 74764

Qy   2607 atatatttgtactatatatttttcatttatatagttttcatttgtatattatatacgtttcatt[illegible] 2666

Db  74765 ATATATTTTATAATATATTATATTTAATATATATTTTATATTATATATTATATATATTTT 74824

Qy   2667 tatattatataattcattatata[illegible] 2726

Db  74825 TATATTATATTATATATATTATATTATATATATTATATATATTATATATAAATTTTATAT 74884

Qy   2727 atgaagcacagagaagttaagtaatttgcgt 2757

Db  74885 ATAATATATAAAGCACATTATAGAATGCCCT 74915


RESULT  13
HS1100E15/c
LOCUS       HS1100E15  103344 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from clone 1100E15 on chromosome X[illegible]22.3
            Contains a pseudogene similar to checkpoint suppressor 1, a
            pseudogene similar to TIP associating protein, a pseudogene similar
            to KIAA0512 and KIAA[illegible], ESTs and STS[illegible] complete sequence.
ACCESSION   AL035551
VERSION     AL035551.6  GI:4826429
KEYWORDS    HTG; pseudogene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103344)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 13, 1999 this sequence version replaced gi:4582[illegible].
            During sequence assembly, data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL, Sw:, SWISSPROT, Tr:, TREMBL
            IMPORTANT: This sequence is not the entire insert of clone 1100E15.
            It may be shorter because we only sequence overlapping sections
```

```
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone 769N13 is at 103245 in this sequence
            The true right end of clone dJ77019 is at 49818 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the Sanger Centre Chromosome X
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            1100E15 is from the library RPCI5 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES              Location/Qualifiers
     source           1   103344
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map="Xq22.1-q22.2"
                      /clone="RP5-1100E15"
                      /clone_lib="RPCI-5"
     repeat_region    141. .153
                      /note="L2 repeat: matches 30.  43 of consensus"
     repeat_region    154. .462
                      /note="AluY repeat: matches 1. .307 of consensus"
     repeat_region    463. .1266
                      /note="L2 repeat: matches 43. .934 of consensus"
     repeat_region    1267. .1819
                      /note="MLT2D repeat: matches 2. .553 of consensus"
     repeat_region    1820. .3602
                      /note="L2 repeat: matches 934. .2344 of consensus"
     repeat_region    3603. .7460
                      /note="L1PA7 repeat: matches 2044. .6066 of consensus"
     repeat_region    7464. .7750
                      /note="L2 repeat: matches 2383. .2746 of consensus"
     repeat_region    7761. .8251
                      /note="L1ME3A repeat: matches 5435. .5940 of consensus"
     repeat_region    8252. .8552
                      /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region    8553. .8727
                      /note="L1ME3A repeat: matches 5264. .5435 of consensus"
     repeat_region    8954. .9521
                      /note="L1ME3A repeat: matches 5555. .6159 of consensus"
     repeat_region    9571. .10926
                      /note="L1MB8 repeat: matches 4851.  6173 of consensus"
     repeat_region    10950. .13284
                      /note="L1MCb repeat: matches 104. .2122 of consensus"
     repeat_region    13349. .16297
                      /note="L1PA8 repeat: matches -1390. .1530 of consensus"
     repeat_region    16298. .16585
                      /note="AluYa8 repeat: matches 1. .285 of consensus"
     repeat_region    16596. .16697
                      /note="L1PA8 repeat: matches 1530. .1639 of consensus"
     repeat_region    16698. .16955
                      /note="L1PA8 repeat: matches 5906. .6163 of consensus"
     repeat_region    16969. .17280
                      /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region    17579. .19867
                      /note="L1 repeat: matches 2125. .4462 of consensus"
     repeat_region    19866. .20970
                      /note="L1M1 repeat: matches 89. .1543 of consensus"
     repeat_region    20972. .21079
                      /note="L1M1 repeat: matches -1389. .-1274 of consensus"
     repeat_region    21090. .21143
                      /note="27 copies 2 mer tt 78% conserved"
     repeat_region    21841. .22259
                      /note="MLI2E repeat: matches 6. .397 of consensus"
     repeat_region    22365. .22407
                      /note="L2 repeat: matches 2700. .2742 of consensus"
     repeat_region    23785. .23859
                      /note="MER5B repeat: matches 1.5. .1?5 of consensus"
     repeat_region    23861. .24023
                      /note="MER63A repeat: matches 1. .1?1 of consensus"
     repeat_region    24024. .24199
                      /note="L1PA2 repeat: matches 5?71. .5946 of consensus"
     repeat_region    24200. .24242
                      /note="MER63A repeat: matches ?1. .?43 of consensus"
     repeat_region    24346. .24517
                      /note="MIR repeat: matches 47. .?1? of consensus"
     repeat_region    25672. .25868
                      /note="L1M4 repeat: matches 4944. .5141 of consensus"
     repeat_region    25869. .26848
                      /note="L1PA2 repeat: matches 5115. .6?93 of consensus"
     gene             27764. .29872
                      /gene="dJ1100E15.1"
     CDS              27764. .29872
                      /gene="dJ1100E15.1"
                      /note="dJ1100E15.1 (similar to checkpoint suppressor 1);
                      match protein: SW:Q07249 PW:Q61572 PW:Q12894 SW:O43638
                      Tr:O13044 Tr:O43821 Tr:O83224; match ESTs: W87246 AA890338
                      AA259017 AA?4?06 AA?03?? W87?52 W86224 H27?70 H82756
                      AA463714 AI?0957 AA?6?96 AI?7?9?? AA?46207 AA283627
                      AA?59??? H4?1?. AA?0?.?? AI?52?86 AA?85?41 AA?85039
                      AA036484 AA481094 AI?0393"
                      /codon_start=1
                      /pseudo
                      /evidence=not_experimental
     repeat_region    28020. .28177
                      /note="44 copies 2 mer aa ?4% conserved"
     repeat_region    30128. .30252
                      /note="L2 repeat: matches 2?94. .3?? of consensus"
     misc_feature     complement(31023. .31267)
                      /note="match: BSS AQ221997"
     repeat_region    31367. .31421
                      /note="MER5B repeat: matches 466. .524 of consensus"
     repeat_region    33278. .33368
                      /note="L1M4 repeat: matches 2??9. .2??? of consensus"
     repeat_region    33364. .33678
                      /note="AluY repeat: matches 1. .3?5 of consensus"
     repeat_region    33679. .34200
                      /note="L1M4 repeat: matches 2140. .2664 of consensus"
     repeat_region    34546. .34749
                      /note="L1PA10 repeat: matches 5954. .6165 of consensus"
     repeat_region    34750. .34828
                      /note="L1PA10 repeat: matches 586?. .5942 of consensus"
     repeat_region    34835. .35067
                      /note="L1M4 repeat: matches 1??4. .2??? of consensus"
     repeat_region    35040. .35346
                      /note="AluSq repeat: matches 1. .?11 of consensus"
     repeat_region    35356. .35561
                      /note="AluJb repeat: matches 67. .?72 of consensus"
     repeat_region    35947. .36365
                      /note="MER21B repeat: matches 1. .?1? of consensus"
     repeat_region    36388. .36973
                      /note="L1M4 repeat: matches 1 ?5. .1??? of consensus"
     repeat_region    37164. .37263
                      /note="40 copies 2 mer tt 88% conserved"
     repeat_region    38589. .38875
                      /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region    39284. .39325
                      /note="21 copies 2 mer ct 88% conserved"
     repeat_region    39411. .39778
                      /note="L1MC5 repeat: matches 1?? . .??32 of consensus"
     repeat_region    39852. .40156
                      /note="L2 repeat: matches 18??. .21?? of consensus"
     repeat_region    40422. .40710
                      /note="AluJb repeat: matches 4. .?8? of consensus"
     repeat_region    40701. .41224
                      /note="262 copies 2 mer ta 57% conserved"
     repeat_region    41284. .41329
                      /note="23 copies 2 mer ta ?6% conserved"
     repeat_region    41711. .42211
```

```
                     /note="LIM4 repeat: matches 2342  2800 of consensus"
     repeat_region   42344..42583
                     /note="I2 repeat: matches 1917  2143 of consensus"
     gene            42914..50062
                     /gene="dJ1100E15.2"
     CDS             join(<42914..42997,43340..43472,44063..44146,44258..44362,
                     44481..44561,44669..44738,44896..44984,45105..45212,
                     45299..45512,47621..47699,47780..47833,47932..48081,
                     48158..48259,48371..48484,48602..48637,49881..50062)
                     /gene="dJ1100E15.2"
                     /note="dJ1100E15.2 (similar to TIP associating protein);
                     match proteins: Tr:Q68985 Tr:Q68984 Tr:Q99793; match
                     cDNAs: AJ122712 AJ122648 U89677 AF031140 AF022130; match
                     ESTs: AA918615 AA418671 AI150012 AA347908 AI328725
                     AI221083 AI208506"
                     /pseudo
                     /codon_start=3
                     /evidence=not_experimental
     repeat_region   43897..43918
                     /note="11 copies 2 mismatches 100% conserved"
     repeat_region   47089..47136
                     /note="24 copies 2 mismatches 77% conserved"
     repeat_region   47396..47435
                     /note="20 copies 2 mismatches 78% conserved"

  Query Match              2.8%;  Score 82.8;  DB 10;  Length 103344;
  Best Local Similarity   52.7%;  Pred. No. 0.00023;
  Matches  275;  Conservative    0;  Mismatches  237;  Indels   10;  Gaps    4;

Qy    2206 aaaggacaaaattgtaaactaagttgaatttttaaatctacttaatttttcaaatgaaag 2265
           ||| | | | |' |||| || |   |' |  |||| ||  || ||    | ||     |
Db   41219 AAAACATATATTTATAAAATATATATTATATAAAAATATATATTTATAAATATATATTAT 41159

Qy    2266 atatatagtcagaggttgaaactttgaaaaca-caaagtgcctctgaatagctatatttt 2324
           || ' |  : |  | ||| | ' |   ||| |   | |  |||  ||||| ||
Db   41158 ATAAAAATATATATTTATAAATATATATTATATAAAAATATATATTTATAAATATATATT 41099

Qy    2325 gtaaacatgaggacatggataagt---gactatataagaggttggattataaatatgtt 2381
           || | |     | ||  |||| |    | | ||||| |   |    ||||||||||| |
Db   41098 ATATAAAAATATATATTTATAAATATATATTATATAAAAATATATATTTATAAATATATA 41039

Qy    2382 tgattttgaaatatgaaatataatgctttaaatcagcctgtttcttaggtttctcaggag 2441
           |' |  |  ||    |  || |  |  ||      | |    ||  | |   |
Db   41038 TTATATAAAAATATATATTTATAAAATATAAATATATTTATAAAATATATATTATATAAA 40979

Qy    2442 accatttttgaaccaggaatgtacattctaaatcaaaatagggtttttaaggtatagtttt 2501
           |  || || |  |  ||| || ||| || | || |||    | | |||  |||| | |
Db   40978 AATATATTTATATAAAAATATATATTATATAAAATATA-TATTATATAATATATATTATA 40920

Qy    2502 tcaaaaacatttgggccataatttatttttctttaatgaatactttttcatttatattat 2561
           |' |' || |    |  || |||  | |   |||  || |  | | | ||||||||
Db   40919 TAAAATATATATTATATAATATATATTATATAAAAATATATATTATATAATATATATTAT 40860

Qy    2562 attttgtgtgtatatattttcattatatagtttcattatatattatatatttgtactat 2621
           ||        ||||| ||   || ||    |    |||| | |||||||| ||  ||
Db   40859 AT-----AAAAATATATATTATATAAATTATATATTATATAAAATATATATTTATATAAT 40805

Qy    2622 atatttttcatttatatagttttcattgtatattatatacctcattatattatataattc 2681
           |||| ||  |  | | |  | |  ||' |  ||||||   || |  |||||   |
Db   40804 ATATATTTATATAAAAATATATATTTATAAAATATATAATATATAAATATATATTTATA 40745

Qy    2682 attatatagttttcatttaatttatttcatttatatactagc 2722
           |  .|.|   |||   | ||| ||| |   | ,| || || ;
Db   40744 AAAATATATATTTTTATAAATATATATTTATATAAAAATATA 40703


RESULT  14
PFMAL4P2
LOCUS       PFMAL4P2   235956 bp    DNA             HTG       11-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL035475
VERSION     AL035475.5  GI:5731898
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 235956)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Jail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 12, 1999 this sequence version replaced gi:5731847.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum/ and also: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..235956
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="4"
BASE COUNT    90265 a  25047 c  22409 g  97435 t    801 others
ORIGIN


  Query Match              2.7%;  Score 81.2;  DB 22;  Length 245456;
  Best Local Similarity   58.8%;  Pred. No. 0.00034;
  Matches  140;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy    2483 gtttttaaggtatagttttttcaaaaacatttgggccataatttatttttctttaatgaat 2542
           ||| || |  || | |||||  |      |||     |      ||||' | | |' | | ;
Db  224375 GTTATTTAATTAAAATTTTTTATTTTTTTTTTTTATTTAATTAAAATTTTTTTTTTTTTTTT 224434

Qy    2543 acttttttcatttatattatattttttctatatatatttttcattatatacgtttattata 2602
           |||||  || ||   || ||||| || | | |  ||| ;' | | ;  ;
Db  224435 TTTTTTTATTTAATTAAATTTTTTTATTTTTTTTTAAATATTTTTTATTTTA 224494

Qy    2603 tattatatatttgtactatatatttttcatttatatagttttcattgtatattattatata 2662
           |  |||||||| || | || |  ||   ||  | |  |||| |'| |;|  |'|;|'
Db  224495 TGATATATATTTTTATTTTAAAAATTTTTTTCTTTTTTTTTTTATTTTATTATATATTA 224554

Qy    2663 tcattatattatataattcattatatagttttcatttaatttatttcatttatatagct 2720
           |   |||||      ||| |  || |' | |  |  |  |  |  |  |  |  |  |
Db  224555 TATATATATATATATTTTTATCATTTTTTTTTTCTTTTTTTTTTTATTTATTTTTT 224612


RESULT  15
AL133216
LOCUS       AL133216   194355 bp    DNA             HTG       11-NOV-1999
DEFINITION  Homo sapiens chromosome 10 clone RP11 291L22, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL133216
VERSION     AL133216.2  GI:6523574
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194355)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
```

```
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 3, 1999 this sequence version replaced gi:6469099.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00010  Length: 5857bp
            Contig_ID: 00018  Length: 1627bp
            Contig_ID: 00114  Length: 1071bp
            Contig_ID: 00147  Length: 7838bp
            Contig_ID: 00164  Length: 3629bp
            Contig_ID: 00189  Length: 1146bp
            Contig_ID: 00210  Length: 13772bp
            Contig_ID: 00219  Length: 2301bp
            Contig_ID: 00236  Length: 2134bp
            Contig_ID: 00274  Length: 4404bp
            Contig_ID: 00281  Length: 6813bp
            Contig_ID: 00286  Length: 3000bp
            Contig_ID: 00350  Length: 3333bp
            Contig_ID: 00378  Length: 958bp
            Contig_ID: 00403  Length: 1635bp
            Contig_ID: 00431  Length: 2303bp
            Contig_ID: 00453  Length: 3733bp
            Contig_ID: 00467  Length: 10762bp
            Contig_ID: 00476  Length: 1256bp
            Contig_ID: 00490  Length: 12060bp
            Contig_ID: 00610  Length: 3175bp
            Contig_ID: 00706  Length: 1021bp
            Contig_ID: 00727  Length: 2610bp
            Contig_ID: 00900  Length: 6257bp
            Contig_ID: 00961  Length: 26808bp
            Contig_ID: 00963  Length: 4889bp
            Contig_ID: 00972  Length: 5458bp
            Contig_ID: 01007  Length: 1106bp
            Contig_ID: 01030  Length: 1015bp
            Contig_ID: 01042  Length: 1748bp
            Contig_ID: 01212  Length: 2177bp
            Contig_ID: 01262  Length: 12938bp
            Contig_ID: 01413  Length: 2304bp
            Contig_ID: 01427  Length: 1600bp
            Contig_ID: 01498  Length: 1839bp
            Contig_ID: 01501  Length: 1678bp.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .194355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-291L22"
                     /clone_lib="RPCI-11.2"
BASE COUNT    50395 a  32567 c  36113 g  47241 t  28039 others
ORIGIN


  Query Match             2.7%;  Score 81.2;  DB 33;  Length 194355;
  Best Local Similarity   63.0%;  Pred. No. 0.00036;
  Matches  145;  Conservative    0;  Mismatches   78;  Indels    7;  Gaps    1;

Qy   2487 ttaaggtatagtttttcaaaaacatttgggccataatttatttttctttaatgaatactt 2546
           |||| | || ||| |   |     ||      || |||| |||  ||| ||   |  ||
Db 137249 TTAATTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAATTTTATTTTATATTTTATT 137308

Qy   2547 tttcatttatattatatttttcgtgtatatattttcattatatacgtttcattatatatt 2606
           |||   || |||| ||| |||  | ||| |  ||| ||| ||||  ||   |||| |  |
Db 137309 IIIIIAIIIAIIIIAIAIIIAIIIIAIIIIAIIIIAIIIIATIIIIATIIIIATIIIIAT 137368

Qy   2607 atatatttgtactatatatttttcatttatatagttttcattgtatattatatacctcat 2666
          || |  || || | |||||||||   ||  | |   ||| ||| ||||||       | ||
```

```
Db 137369 ATTTTATTTTATTTTATATTTTATTTTATTTTATATTTTATTTTATATT-------TTAT 137421

Qy   2667 tatattatatgattcattatatagttttcatttaatttattcgttttat 2716
          | |||| |||| || || |
Db 137422 TTIATTTTATATTTTATTTTATTTTTTTTATTTATTTTATTTTATTTTAT 137471


Search completed: February  9, 2000, 08:54:41
Job time: 16195 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 02:53:51 ; Search time 159.56 Seconds
(without alignments)
2153.245 Million cell updates/sec

Title: US-09-174-937-7
Perfect score: 2966
Sequence: 1 gtcgaccacgcgtccgggc.....aaaaaaaaaagggcggccgc 2966

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/1/ina/7A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/7B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/7C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/7E_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	75.2	2.5	19124	4	US-08-487-826B-13	Sequence 13, Appl
c 2	70.4	2.4	240	1	US-08-[illegible] 417-6	Sequence 6, Appli
3	68.4	2.3	7218	1	US-08-[illegible] 463-14	Sequence 14, Appl
c 4	68.2	2.3	6243	4	US-09-[illegible] 075-1	Sequence 1, Appli
c 5	67.4	2.3	2447	4	US-09-[illegible] 969-14	Sequence 14, Appl
c 6	59.4	2.0	5852	1	US-07-[illegible] 106-2	Sequence 2, Appli
c 7	58	2.0	3095	6	5231169-1	Patent No. 5231169
c 8	56.4	1.9	1183	4	US-08-[illegible] 937-10	Sequence 10, Appl
9	56.2	1.9	1422	2	US-08-[illegible] 704-5	Sequence 5, Appli
10	55.6	1.9	19124	4	US-08-487-826B-13	Sequence 13, Appl
11	54.8	1.8	665	4	US-08-[illegible] 795A-36	Sequence 36, Appl
12	54.2	1.8	5852	1	US-07-[illegible] 106-2	Sequence 2, Appli
13	52.2	1.8	731	2	US-08-[illegible] 405A-2	Sequence 2, Appli
14	52.2	1.8	8920	3	US-08-[illegible] 855A-1	Sequence 1, Appli
c 15	52.2	1.8	8920	3	US-08-[illegible] 855A-1	Sequence 1, Appli
c 16	52	1.8	8595	1	US-08-[illegible] 096-3	Sequence 3, Appli
17	51.8	1.7	6243	4	US-09-[illegible] 075-1	Sequence 1, Appli
18	51	1.7	3001	2	US-08-[illegible] 100-1	Sequence 1, Appli
19	51	1.7	2993	2	US-08-[illegible] 100-2	Sequence 2, Appli
20	51	1.7	643	2	US-08-[illegible] 100-7	Sequence 7, Appli
21	51	1.7	602	2	US-08-[illegible] 100-8	Sequence 8, Appli
c 22	51	1.7	2993	2	US-08-[illegible] 100-10	Sequence 10, Appl
c 23	50.4	1.7	5181	2	US-08-[illegible] 073-10	Sequence 10, Appl
c 24	50.2	1.7	3[illegible]	2	US-08-[illegible] 100-9	Sequence 9, Appli
c 25	50.2	1.7	642	2	US-08-[illegible] 100-13	Sequence 13, Appl
c 26	50.2	1.7	2059	4	US-09-[illegible] 991-1	Sequence 1, Appli
c 27	50.2	1.7	665	4	US-08-[illegible] 795A-36	Sequence 36, Appl
28	49.6	1.7	7904	1	US-08-[illegible] 239B-1	Sequence 1, Appli
29	49.6	1.7	7904	1	US-08-[illegible] 239B-2	Sequence 2, Appli
30	49.6	1.7	7904	1	US-08-[illegible] 005-1	Sequence 1, Appli
c 31	49.4	1.7	4970	2	US-08-[illegible] 100-14	Sequence 14, Appl
c 32	49.4	1.7	473	2	US-08-[illegible] 100-16	Sequence 16, Appl
33	49.4	1.7	4970	2	US-08-[illegible] 100-20	Sequence 20, Appl
c 34	49.4	1.7	19557	5	[illegible]	[illegible]
35	49	1.7	4298	4	[illegible]	[illegible]
c 36	48.8	1.6	1511	1	[illegible]	[illegible]
c 37	48.8	1.6	660	1	[illegible]	[illegible]
38	48.8	1.6	1002	2	[illegible]	[illegible]
39	48.8	1.6	1511	2	[illegible]	[illegible]
c 40	48.8	1.6	660	2	[illegible]	[illegible]
c 41	48.8	1.6	1511	4	[illegible]	[illegible]
c 42	48.8	1.6	660	4	[illegible]	[illegible]
43	48.8	1.6	1002	5	[illegible]	[illegible]
44	48.8	1.6	1000	5	[illegible]	[illegible]
c 45	48.6	1.6	1939	1	[illegible]	[illegible]

ALIGNMENTS

RESULT 1
US-08-487-826B-13/c
; Sequence 13, Application US/[illegible]
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: S, Xiaohaun
; APPLICANT: Williams, Thomas E.
; TITLE OF INVENTION: BINDING [illegible]
; TITLE OF INVENTION: AND [illegible] PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive, 16th Fl
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.3 ICP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 2.5%; Score 75.2; [illegible]
Best Local Similarity 48.6%; Pred. No. [illegible]
Matches 241; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible]

Qy 2233 attttaaatctacttaatttt[illegible] 2292

Db 16195 ATATGTGTATGTATATATGTATTTA[illegible] 16126

```
Qy   2293 aaaaaaaaaatgcctctgaatagctatatttctaaaacatgaggacatggataagtgact 2352
          ||  | | |  || |||| | || | ||  | |  || | | |||  |
Db  16125 TGTAGTTATGGGTATATATATATATATATATAATATATATATGTATGTGTAAATGTGTAT 16066

Qy   2353 gtataaaaggtttggattataaatatgtttgatttctgaagtatgaaaataaatgcttgaa 2412
          |||    ||||     |||| |||| | ||  |    || | |   ||  |    ||
Db  16065 GTAGTTATGGGTATATATATATATATATATATGTATGTGTAAATGTTGTATGTAGTTATGGG 16006

Qy   2413 agcagcctgttttcttaggttttctcaagcgaccattttgaaccaggaatgtacattctaaa 2472
          |  |   |  ||||| | |    |   |    |  | | ||  || |  |
Db  16005 IATATATATATATATATATGIATGIATATGIGCA:GTAATTAATTTTATTTTATTTTTTTTA 15946

Qy   2473 cnnnatgggggttttttaaggtatagttttttcaaaaacatttgggcca·····taattt 2525
          |||    |||||| |  | |  || ||  | | ||||    |      | ||||
Db  15945 TTTAATTAAATTTTTTTATTTTTTTTTTATTTCATTAAATTTTTTTAATTTTTTTTTATTT 15886

Qy   2526 attttcttttaatgaatacttttttcatttatattatattttcgtgtatatattttcatt 2585
          | ||  |||  | | |  || ||  ||| | |  |||     | | ||| | ||  ||
Db  15885 AATTAAATTTTTTTATTTATTATTATTTTTTTTIATTTAATTAAATTTTTTTATTTATTATTT 15826

Qy   2586 atatacgtttcattatatattatatatttgtactatatatttttcatttatatagttttc 2645
          | |   ||| | || || || | |||     ||||||||||  ||||    ||||
Db  15825 TTTTTTATTTAAATAAATTTTTTTTATTTIATGTATATATTTTTTTTTTAACATTTTTTT 15766

Qy   2646 attgtatattatatacctcattatattatataattcattatatagttttcatttaattta 2705
          | | ||  ||   |  | ||     |   |  || | |||'|  ||||| |||  |||
Db  15765 AATTTTTTTTTATTTTATGATATATATTTTIATTTAATATATTTTTTTCTTTTTTTTTT 15706

Qy   2706 tttcattttataacta 2721
          |  |  | |  |||  ||
Db  15705 GTTTTTATGATATATA 15690


RESULT   2
US-08-628-417-6/?
; Sequence 6, Application US/08628417
; Patent No. 5627054
;  GENERAL INFORMATION:
;    APPLICANT:  GILLESPIE, DAVID
;    TITLE OF INVENTION:  COMPETITOR PRIMER ASYMMETRIC
;    TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  U.S. ARMY CHEMICAL AND BIOLOGICAL
;      ADDRESSEE:  DEFENSE COMMAND
;      STREET:  OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
;      CITY:  ABERDEEN PROVING GROUND
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21010-5423
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,417
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIFFONI, ULYSSES J
;      REGISTRATION NUMBER:  39,908
;      REFERENCE/DOCKET NUMBER:  DAM 396-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  410-671-1158
;      TELEFAX:  410-671-2534
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligodeoxynucleotide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  YES
US-08-628-417-6


  Query Match              2.4%;  Score 70.4;  DB 1;  Length 24 ;
  Best Local Similarity   58.7%;  Pred. No. 2.6e-07;
  Matches  122;  Conservative    0;  Mismatches   86;  Indels     7;  Gaps    0;

Qy   2523 tttatttttcttttaatgaatacttttttcatttatattatattttttcatatatatttttt 2582
          ||| ||||  |||  |   |  |  |||||  ||| | || | |||||  | |  |  | |
Db    237 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 178

Qy   2583 attatatacgtttcattatatattatatatttgtactatatatttttcatttatatatt 2642
          || | |   |||  || | | |  |  |  | | |  | |  |  |  |  |  |  |
Db    177 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 118

Qy   2643 ttcattgtatattatatacctcattatattatataattcattatatagttttcatttaa 2702
          || || | || | |   | |  | | | | | | |  |  |  |  |  |  |  |  |
Db    117 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 58

Qy   2703 ttatttcattttataactagaacactga 2730
          || |||  |||| |   ||| |  | | |
Db     57 TTTTTTTTTTTTTTTTTTAGTAAAATTA 30


RESULT   3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
```

```
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match              2.3%;  Score 60.4;  DB 1;  Length 7218;
  Best Local Similarity    2.7%;  Pred. No. 2.1e-06;
  Matches    9;  Conservative  210;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     758 gagcttgccaatttcttgtataagttcccactatctgcaatgttcttttccccagatcct 817
           ||||||| | ::::::  :  :::::  :  :::  :  :::::::::::  ::::
Db    1056 GAGCTTGTGGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1115

Qy     818 tacgtgactgactccttctctttattcagatgttaatcagatgtcacctcctcagagaag 877
           :  :  ::::::  :  ::::  :  ::  :  :  :  :::::::
Db    1116 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1175

Qy     878 tcttccttgacctctgtaatcaaagatgcttcccacttcccgactcccaccacctagtca 937
           :::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    1176 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1235

Qy     938 ctctctgtcccagtgttcatttttatttccacgtatcactaattcaaattgtattattta 997
           :::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    1236 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1295

Qy     998 actatttgtttgcatctttattgtctttctcttaccactcgaatgtatgctccacaggag 1057
           ::  :..  :  :  :::::::  ::  ::  :  :  :  :::  :
Db    1296 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1355

Qy    1058 caaggactttatatttgatcctgcgtctcc 1087
           :  :  :::  :  :::  :::::  :  :
Db    1356 YYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1385


RESULT   4
US-09-056-075-1/c
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  360295.97038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3770..4013
;      OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;      OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1


  Query Match              2.3%;  Score 60.2;  DB 1;  Length 6243;
  Best Local Similarity   56.9%;  Pred. No. 2.1e-06;
  Matches  145;  Conservative    0;  Mismatches  110;  Indels    2;  Gaps    1;

Qy    2465 atttaaaaaaaaaatagggttttttaggttttaggtttttaaaaaaaaaaaaaaaaaaaa 2524
           ||   ||| |||||  ||  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db    1507 ATATGAAAAAAAAATGGGCTTTTATAAAGGTTCATTTTTTATATTTTTTTTAAAGATT 1448

Qy    2525 tatttttttttaatgaatgttttttaattatattatttttatgtatatttttttttaaaat 2584
           |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    1447 ATATAATATAAAAAAAATTTTTTCAAACTTTAAATAAAAAATATTTTTATATTTTTTTAT 1388

Qy    2585 tatatacgtttcattatatattttatatttgtactatttatttttcc--tttatatgatt 2642
           |  |  |   |||  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    1387 TTTTTTATTTTTATATTTTTTTATTTTTTTATTTTTATATTTTTTTATTTTTTATATTTT 1328

Qy    2643 ttcattgttatattatatatttttttatttatatatttttttatatttttttttttttt 2702
           ||  ||  ||||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    1327 TTATTTTTATATTTTTTTATTTTTATATTTTTTTATTTTTTTATTTTTTTATTTTTTAT 1268

Qy    2703 ttatttcatttttata 2717
           ||  |  |  ||||  ||
Db    1267 TTTTTTATTTTTTTA 1253


RESULT   5
US-09-114-969-14/c
; Sequence 14, Application US/09114969
; Patent No. 5965397
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John M.
;    APPLICANT:  LaVallie, Edward R.
;    APPLICANT:  Racie, Lisa A.
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Spaulding, Vikki
;    APPLICANT:  Agostino, Michael J.
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/114,969
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.
;      REGISTRATION NUMBER:  41,329
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  14:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-014-969-14


  Query Match              2.3%;  Score 67.4;  DB 4;  Length 2447;
  Best Local Similarity   57.5%;  Pred. No. 2.5e-06;
  Matches  119;  Conservative    1;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    2523 tttatttttcttttaatgaatactttttcatttatattatatttttcgtgtatatatttttc 2582
           ||| |||| |||| |  |  ||||| ||| | || | ||||| | | | | ||||
Db    2440 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2381

Qy    2583 attatatacattttcattatatattatatatttctactatatttttcatttatatagtt 2642
           || | |   ||| || | | || | | ||| |  | | | ||||  ||| | |  ||
Db    2380 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2321

Qy    2643 ttcattgtatattatatacctcattatattatatatttcattatatagttttcatttaat 2702
           ||  || | | || | |   | || | | | | ||  || | |  |||| ||| |
Db    2320 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2261

Qy    2703 ttatttcatttttataactagaacaatg 2731
           ||  |||  |||| |  ||  | ||| |
Db    2260 TTTTTTTTTTTTTTTTTTTTTTTTAAAAGG 2234


RESULT   6
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & Nc. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  13220625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/0053.
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-1002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match              2.0%;  Score 59.4;  DB 1;  Length 5852;
  Best Local Similarity   54.5%;  Pred. No. 0.00019;
  Matches  140;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy    2490 aggtatagtttttcaaaaacatttgggccatatttatttttttttaatgaatatttttt 2549
           ||| | | |  | ||||  |     |   |||| | | | | | ||||  |
Db    5851 AGGAAGAATATTACAAAGTAACAATACTACTAATAAAACATCTTTTAAATTTATTTATAT 5792

Qy    2550 catttatattatatttttcgtgtatatatttttcattatatagtttcattatatatttata 2609
           ||||| || |   | |||  | ||| |  | ||  | ||| | |  |||| | |  |||
Db    5791 CATTTTTAATGGTTATTTAATTTATTTAATTATTTCTTATTTTTATATTTTTTATATATAT 5732

Qy    2610 tatttgtactatatattttcatttatatagtttcatttatattatatacttcattat 2669
           |||||||  | | | |||||  ||    |  | ||  ||  | ||| | | |  |||
Db    5731 TATTTGTTGTTTGTTTTTTAGTTTATTTCTATTTTTTATTTATAAAATTAATAATTAA 5672

Qy    2670 attatataattcattatatagtttcatttaatttatttcatttttataaat 2720
           ||| || |||  || | | |      |   ||   |||| | || | ||  | || ||
Db    5671 ATTTTA-AATAAATAAAAAAAAAAAAAAAAAAAAAATTTAAATTAAAAAAAAAAAAAAA 5613

Qy    2730 aagcacagaggaagttaa 2746
           ||  | | | | | | ||
Db    5612 AAAAAAAAAAAAATAAA 5596


RESULT   7
5231168-1/c
;Patent No. 5231168
;    APPLICANT: DZIEGIEL, MORTEN,BORRE, MARTIN,JEPSEN, SIMON;
;VUUST, JENS;RIENECK, KLAUS,WIND, ANNETTE,JAKOBSEN, PALLE H.
;    TITLE OF INVENTION, MALARIA ANTIGEN
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/409,658
;      FILING DATE: 18-SEP-1989
;SEQ ID NO:1:
;      LENGTH: 3095
5231168-1


  Query Match              2.0%;  Score 58;  DB 6;  Length 3095;
  Best Local Similarity   50.0%;  Pred. No. 0.00032;
  Matches  145;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    0;

Qy    2491 tttcttaagcgagaaatttgaa                                        2490
           ||| | ||  |   | ||| |      ||  || | | | |||                 |
Db    2636 TTTATTACATGTATACTTTCAGATTATAAAATATACTCTAAATATTTTTTATATTTTATTT 2577

Qy    2491 ggtatagttttttaaaaaacatttgggccatatttatttttttttaatgaatatttttt 2550
           ||||| ||||  || |  | |  ' '  '                              '
Db    2576 ATTATATTTTTAGAATATTAATACAACTAAGAAAACAAATTATTTATTTTTTTTTTTTTT 2517

Qy    2551 atttatattatatttttcgtgtatatatttttcattatatagtttcatttatatatttata 2610
           | ||| |||  |  | |  '  '  '  '  '  '  '  '  '  '  '  '  '
Db    2516 TATAATAATATACAATGAATTTATTTCTATATATACACAAATCCTATTATTTTTTTTTTT 2457

Qy    2611 atttgtactatatattttcatttatatagtttcatttatattatatacttcattat 2670
           | |    ||| ||||  | | | ||||    || |||  |||  | | |  |  |
Db    2456 AATATATATATTTATTAATAAATAATATTAGTTACATATATATAAAAAAAAAAATTTTAT 2397

Qy    2671 ttatataattcattatatagtttcatttaatttatttcatttttataaat 2720
           | ||||  ||  || | |  ||||  |||  |||  || | | |  ||| |
```

```
Db   2396 TAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTAATTATTTAAAT 2347


RESULT   8
US-08-705-937-10/c
  Sequence 10, Application US/08705937
; Patent No. 5981841
;  GENERAL INFORMATION:
;    APPLICANT:  Santino, Colleen G.
;    APPLICANT:  Conner, Timothy W.
;    TITLE OF INVENTION:  EARLY SEED 5' REGULATORY SEQUENCE
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carmen Rodriguez, Paralegal, Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,937
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nicolas G. Barcoukas,
;      REGISTRATION NUMBER:  29,922
;      REFERENCE/DOCKET NUMBER:  MOBT.019 (39-2(10694)A
;      NAME:  Janelle D. Waack,
;      REGISTRATION NUMBER:  36,300
;      REFERENCE/DOCKET NUMBER:  MOBT.019 (39-2(10694)A
;      NAME:  Barbara S. Kitchell
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  MOBT.019 (39-2(10694)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 787-1400
;      TELEFAX:  (713) 789-2679
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-705-937-10


  Query Match             1.9%;  Score 56.4;  DB 4:  Length 1183;
  Best Local Similarity   49.8%;  Pred. No. 0.00053;
  Matches  132;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy   2484 tttttaaggtatagtttttcaaaaacatttgggccataatttatttttctttttaatgaata 2543
          |||| | | ||  |||  ||| |||      |  || | |||||  |  |||
Db    535 TTTTAATTGATTAAANATTTATTAAATATTATAAATTTTTAAGGTTATATTTTATTTNATA 476

Qy   2544 ctttttcatttatattatatttttcgtgtatatattttcattatatacgtttcattatat 2603
            ||||   , ||  | || || |   | | | .|  |||| |  |||      | |||||
Db    475 ATTTTCTCTCTAGCTAATTTTATATTTTTTTTTAAATTCAATCAATANCAAAGAATATAT 416

Qy   2604 attatatatttgtactatatatttttcatttatatagttttcattgtatattatatacct 2663
               || |  |  |||||||||| |      | |||    || | | |        ||
Db    415 TAANNANATGTTAAAAATATATTTTTAACCCTCCTCGTTCAATTTTTTTTTATGGATGCT 356

Qy   2664 cattatattatataattcattatatagttttgatttaatttatttcattttataactaga 2723
            |  ||   |  || |  | | |  |  ||||| |||| ||   | |
Db    355 AAANAGTGTAATATTTATGTTTCAGTATGTCCTTAAATAATATTTCTTTTTGTATGGATA 296

Qy   2724 acaatgaagcacagaagaagttaagt 2748
          | | | |  |  | ||  ||||
Db    295 ACTATTATATACTAAGGAGTGAAGT 271


RESULT   9
US-08-319-704-5
  Sequence 5, Application US/08319704
; Patent No. 5814617
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard C.
;    APPLICANT:  Doolan, Denise L.
;    TITLE OF INVENTION:  Protective 17 Kda Malaria Hepatic Anti
;    TITLE OF INVENTION:  Erythrocytic Stage Immunogen and DNA
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Naval Medical R & D Command
;      STREET:  Bldg 1, T-12, 8901 Wisconsin Avenue
;      CITY:  Bethesda
;      STATE:  Maryland
;      COUNTRY:  U.S.A
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/319,704
;      FILING DATE:  07-OCT-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. David Spevack
;      REGISTRATION NUMBER:  24,749
;      REFERENCE/DOCKET NUMBER:  PD-2.0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6754
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1422 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-319-704-5


  Query Match             1.9%;  Score 56.2;  DB 4:  Length 1422;
  Best Local Similarity   49.8%;  Pred. No. 0.00064;
  Matches  142;  Conservative    0;  Mismatches  [illegible]  Indels    0;  Gaps    0;

Qy   2469 taaaaaaaaataggggttttttaatatttttttttaaaa attgaa  aaaattttt 2528
          |  | | |||  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      3 TAAATCATATATGATATATTATATTTTATATATGGTATTATTTTTTAAAAAAAATTATA 62

Qy   2529 ttttttaatgaatacttttttaaattatttatttttaaaaaaaaaaaaaaaaaaaaaaa 2588
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     63 TTTTAATATAGTTAGCGAGATTTAAAAAATAATTATTTTATTTTATAAATTTATATTATA 122

Qy   2589 tacgtttcattatatattatatttttgttatatatttttttgattgtgtttttttttt 2648
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    123 GATTATTATCTCTATATAGAAAAAATATTAAAAAAATATTTTTATAATTTTTTATAAATT 182

Qy   2649 gtatattatatacctcattatattatatattatatatatttttttttttaaaaaaaaaaa 2708
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    183 TTAATTTTTATTATTAATTTTGTATAAAAATAATATTTATTTTTTTTATAAATTTAAATT 242

Qy   2709 cattttataactagaaacaatgaagcaacagcaagttaagtaaaaa 2754
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    243 TTTAATAAAATATTATTAATTTTATTTATTTAATTAAATTTTTTTT 288
```

```
RESULT  10
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  MIN121.1-1CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             1.9%;  Score 55.6;  DB 4;  Length 19124;
  Best Local Similarity   51.3%;  Pred. No. 0.0019;
  Matches  155;  Conservative    0;  Mismatches  144;  Indels    3;  Gaps    1;

Qy    2455 agaaatgtacattctaaacaaaatagggttttttaaggtatagttttttcaaaaacatttg 2514
           |  ||||  |   |  |||  ||||   |  ||   |||    |  |  |  |  | || |
Db   17538 ATGAATATTGTTAATAATAAAAAGGAGATTTTTGAAGAGGAATATCCTATATCAGATATA 17597

Qy    2515 ggccataatttatttttcttttaatgaaat---actttttcatttatattatatttttcgtg 2571
           |  ''  |  |  ||  ''|  |  |    |  | |||| ||||||  | |  ||   ||
Db   17598 TGGAATATCTAAAATTAATATACTTTTTTTGTGTGTGTCATATATATTTTGTATTATTTG 17657

Qy    2572 tatatattttcattatatacgtttcattatatattatatatttgtactatatatttttca 2631
           ''|''  |  |  |''| |||   |''|  ||  |   ||   |  |||  |  ||||| |  |    |
Db   17658 TATATGTTTTTATTTTATTTATTTATTTATTTATTTATTGTTTTTGGTATATTTGTAAAA 17717

Qy    2632 tttatatagttttcattgtatattatatacc'cattatattatataattcattatatagt 2691
             ''|     |  |   |    |  |||       ||  || '' |  |    ||| || |  |
Db   17718 AATATGTTTTTGTTTATAATCATATTATTATATTTTTAATAATTTGCAACATGATTTTTT 17777

Qy    2692 tttcatttaatttatttcatttttataactagaacaatgaagcacagagaagttaagtaat 2751
           |||   ''  '|  |  |  |''  |  |    |  | |  |  | | | | ||    || |
Db   17778 TTTTTCTTTCTTATTGTGTAATTTTTTTCATAATATTTATATATATATATGTATTTTATT 17837
```

```
Qy    2752 tt 2753
           ||
Db   17838 TT 17839


RESULT  11
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
;  GENERAL INFORMATION:
;    APPLICANT:  Delcuve, Genevieve
;    APPLICANT:  Awang, Gregor
;    TITLE OF INVENTION:  Recombinant DNA Molecules and Expression
;    TITLE OF INVENTION:  Vectors for Tissue Plasminogen Activator
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BERESKIN & PARR
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,795A
;      FILING DATE:  27-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7841-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  665 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  Rh 32
US-08-883-795A-36


  Query Match             1.8%;  Score 54.8;  DB 4;  Length 665;
  Best Local Similarity   52.5%;  Pred. No. 0.001;
  Matches  146;  Conservative    0;  Mismatches  127;  Indels    5;  Gaps    1;

Qy    2481 gggtttttaaggtatagttttttcaaaaacatttgggccataatttatttt'''''''''' 2540
           |  |||||||  |  | | ||| | ||  |          |||||  |  |  |  |  |
Db      18 GTGTTTTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAAATATTTATAATTA 77

Qy    2541 atacttttttcatttatattatatttttcgtgtatatatttttcatttatatatttttttta 2600
           |   ||||  | ||| | ||| | |          | | |   | | ||| |  |  |  |
Db      78 AATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTA 137

Qy    2601 tatattatatatttgtactatatatttttcatttatatagttttcattatatattatata 2660
           |||||  |||| ||  |  ||| |||  ||  | |  |  |          |' |
Db    138 AATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTA 197

Qy    2661 cctcatta-----tattatataatttcattatatagttttcatttaatttat'''''''' 2715
             |  ||      ||  ||||   |  '|||  |'|  |||   |  ,'|   '' |  |
Db    198 AATATTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTA 257

Qy    2716 taactagaacaatgaagcacagagaagttaagtaattt 2753
```

```
Db    258 AATATTTTATAATTAAAATGTTTATAATTAAATATTTT 295


RESULT  12
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  PICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             1.8%;  Score 54.2;  DB 1;  Length 5852;
  Best Local Similarity   51.5%;  Pred. No. 0.0027;
  Matches  152;  Conservative    0;  Mismatches  138;  Indels    5;  Gaps    1;

Qy    2421 gtttcttaggttttctcacgcgaccattttgaaccaggaatgtacattctaaacaaaatag 2480
Db    2071 GTTTTTTTTGATTTTTTTTTTAATAAATTTAAAATTTTATTCTCTATCTAATTATACCTT 2130

Qy    2481 gggtttttaaggtatagttttttcaaaaacatttgggccataatttatttttcttttaatga 2540
Db    2131 ATTTATAAATATTGGATAATATATCAAATATTTATCAGTTTTGGCATGACAATTTTAATT 2190

Qy    2541 atacttttttcatttatattatatttttcgtgtatatatttttcattatatacgtttcatta 2600
Db    2191 ATATTTATTTTTTGATTAATTTTTTTTTTTTTTTTTTTTTAAAAATTCITTTTTTTTTTT 2250


Qy    2601 tatattatatatttgtactatatatttttcatttatatagttttcattgtatattatatc 2660
Db    2251 TTTATTTTTAATTTTTAATTTTTATTTTCCGACATTTCATTTTTMTTTTATTTA 2310

Qy    2661 cctcat-----tatattatatgattcattgtatgttttgattgattttattttg 2711
Db    2311 TTGTAAATTCATTTTATTTATTTTTAATTAAATAGTTTTTTTTTAATTTTATTTA 2365


RESULT  13
US-08-451-405A-2
; Sequence 2, Application US/08451405A
; Patent No. 5736358
;  GENERAL INFORMATION:
;    APPLICANT:  FASEL, NICOLAS JOSEPH
;    APPLICANT:  REYMOND, CHRISTOPHE DOMINIQUE
;    TITLE OF INVENTION:  DICTYOSTELIS EXPRESSION VECTOR AND
;    TITLE OF INVENTION:  METHOD FOR EXPRESSING A HETEROLOGOUS PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE WEBB LAW FIRM
;      STREET:  700 KOPPERS BUILDING, 436 SEVENTH AVENUE
;      CITY:  PITTSBURGH
;      STATE:  PENNSYLVANIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  15219-1818
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" FLOPPY DISK
;      COMPUTER:  Midwest Micro 486-SX
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WORDPERFECT 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/451,405A
;      FILING DATE:  26-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/895,273
;      FILING DATE:  15-JAN-1993
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  731
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  UNKNOWN
US-08-451-405A-2


  Query Match             1.8%;  Score 52.2;  DB 2;  Length 731;
  Best Local Similarity   52.5%;  Pred. No. 0.0039;
  Matches  114;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy    2484 tttttaggtatagttttttcaaaaaaaattgtttgtatttatttttatttaaaaaaaaa 2543
Db     350 TTTTTAAAATATTATTATATATTTTTAATTTTTTCATAATCTTTTTTTTCATATTTATT 409

Qy    2544 ctttttcatttatattatattttttgtgtatatatttttcattatatatgtttttttatat 2603
Db     410 ATCTTACAAAGGTTTAAAATTTTAATTTTTTTTATTTATTATTATTTTTTTAAATAAAT 469

Qy    2604 attatatatttgtactatatatttttttttttttttttttttttttttttttttttttt 2663
Db     470 TTTTTTTAATTTTTTAATTTTTTTTTTTTTTTTACTATTAATTTTTTTTTTTTTTTTT 529

Qy    2664 cattatattatataattcattatatatttttattta 2700
Db     530 AACATTTATTATTTATTTATTTTATATATTTAATTA 566


RESULT  14
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5649573
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER  29,009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-855A-1

  Query Match             1.8%;  Score 52.2;  DB 3;  Length 8920;
  Best Local Similarity   52.0%;  Pred. No. 0.0084;
  Matches  117;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy   2522 atttatttttttttaatgaatatttttttattttatattatttttcgtgtatatatttt 2581
          ||||||||  ||      | |    ||| | |  || || || |||    || |||| |
Db   8542 AITTATTTATTTAAITAIATTTACGTTTAAAIAITAATAAATGTTTTTATTAAATATGAT 8601

Qy   2582 cattatatacgtttcattatatattatatatttgtactatatatttttcatttatatagt 2641
           ||   |   || |||   | | || || |    |    ||||| |   | || |||
Db   8602 CATTAATTTATATTGATTTATTTTTTTATAAATTTTTGTTATATATACAAATTTTATTTA 8661

Qy   2642 tttcattgtatattatatacctcattatattatataattcattatatagttttcatttaa 2701
          ||   |  ||| |   |  |   ||  | ||  | ||    |||   || | |||| |
Db   8662 TTTATTATATATTATAAATTAAAATGTTTTTTTAAATTTATAAAATAAATTTTATAATTTTA 8721

Qy   2702 tttatttcatttttataactagaacaatgaaaaacagagaagttaa 2746
          | | || |||   | |   | || | |  || | | | | | ||||
Db   8722 ATAAATTTATTAATTATAAAAAAAATAAAAATATATAAACATTAA 8766

RESULT  15
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29,009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-855A-1

  Query Match             1.8%;  Score 52.2;  DB 3;  Length 8920;
  Best Local Similarity   50.2%;  Pred. No. 0.0084;
  Matches  155;  Conservative    0;  Mismatches  153;  Indels    1;  Gaps    1;

Qy   2445 attttgaaccaggaatgtacattctaaacaaaatagggtttttaaagtataattttttca 2504
          | ||| ||  |  | | |    | | || | ||||   | ||| ||  |  |
Db    867 ACTTTTAATAACAACTTTGTGATGTTAAAAGAATAAAACTGTTTAAGAGTTATTATTCAT 808

Qy   2505 aaaacat-ttgggccataatttatttttttttaatgaatagttttttcatttatattatt 2563
          | || ||          || || | | |  | |  | |  |  |  |  |  |  |
Db    807 AGAATATCCCCAATAATTATATATATATATATTTATATATATATGTATATTATTTTT 748

Qy   2564 ttttgtgtatatattttcattatttagtttattatatattatttatttttatttttt 2623
          | | | | ||||  |||  | | | |  |||  | |  | |  |  |  |  |
Db    747 CCCATTTTTCTTTTTTTTTTTATACATTTATATTTATGTTTAAAATATTTATAAATTTA 688

Qy   2624 atttttcatttatatagttttcattgtatattatatacctcattatatttatattttt 2683
          | | ||  |  ||  ||   ||  |  || | |  |  || | |  |  |  |
Db    687 CATATACAAGTTCATTTTTCATATTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTT 628

Qy   2684 tatatagttttcatttaatttatttcattttataactagaacaatgaaaaacagagaagtt 2743
          | | |  |||| | |||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    627 TTTTTTTTTTAAATTATTATAAATTAATAAAAAAAAATAAAAATATATTTTTTTTT 568

Qy   2744 taagtaatt 2752
          ||  ||||
Db    567 GAATAAATT 559

Search completed:  February 10, 2000, 03:01:11
Job time:  1649 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2000, 19:33:19 ; Search time 3770.17 Seconds
(without alignments)
2870.322 Million cell updates/sec

Title: US-09-174-937-7
Perfect score: 2966
Sequence: 1 gtcgaccaacgcgtccgggc.....aaaaaaaaaaggcgcggcgc 2966

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831981 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to that shown for the result listed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	401	15.0	497	36	AA841789	AA841789 ak09a04.s
2	457.4	15.4	570	104	AJ560224	AJ560224 HS_5340_B
3	371.4	12.5	424	104	AJ561104	AJ561104 HS_5340_B
4	301.2	10.2	302	24	BE156...	BE156... HUMT1BG12B
5	258.4	8.7	275	84	AA814...69	AA814...69 nm40g09.s
6	184.8	6.2	429	20	AJ140851	AJ140851 HS_5207_A
7	179.4	6.0	236	31	AA248...	AA248... EST1097...

c	8	106.8	3.6	565	104	AQ533091	AQ533091 RPCI-11-3
c	9	92.8	3.1	1101	79	CNS00EVL	AL069706 Drosophil
c	10	91.8	3.1	1101	79	CNS00Z6Z	AL097301 Drosophil
c	11	91	3.1	1101	79	CNS00LI2	AL078714 Drosophil
c	12	88	3.0	1101	79	CNS00396	AL063921 Drosophil
c	13	87.6	3.0	1101	79	CNS00211	AL061936 Drosophil
c	14	85.2	2.9	734	79	CNS010MF	AL099163 Drosophil
	15	85.2	2.9	1101	80	CNS017V2	AL108536 Drosophil
	16	84.8	2.9	1101	79	CNS00EVL	AL069706 Drosophil
c	17	84.2	2.9	1101	80	CNS017KF	AL108152 Drosophil
c	18	83.6	2.8	996	79	CNS005TH	AL071063 Drosophil
	19	83	2.8	1101	79	CNS00EIW	AL069847 Drosophil
	20	83	2.8	1101	80	CNS01807	AL108721 Drosophil
c	21	82.8	2.8	1069	79	CNS00FTY	AL069946 Drosophil
c	22	82.6	2.8	842	79	CNS0110N	AL099305 Drosophil
c	23	82.6	2.8	806	79	CNS0090W	AL052249 Drosophil
	24	82.4	2.8	1101	79	CNS00396	AL063921 Drosophil
c	25	82	2.8	905	79	CNS00KHX	AL077798 Drosophil
	26	81.8	2.8	922	79	CNS0090W	AL052618 Drosophil
	27	81.6	2.8	1009	79	CNS00787	AL057153 Drosophil
c	28	81	2.7	1101	79	CNS00B01	AL057419 Drosophil
c	29	81	2.7	467	80	CNS01838	AL109298 Drosophil
c	30	80.8	2.7	987	80	CNS014P5	AL104456 Drosophil
c	31	80.2	2.7	843	79	CNS0091	AL053009 Drosophil
	32	80.2	2.7	1101	80	CNS017Y1	AL108642 Drosophil
	33	80	2.7	1101	79	CNS001FB	AL060732 Drosophil
	34	80	2.7	990	79	CNS00601	AL065624 Drosophil
c	35	80	2.7	614	80	CNS0152H	AL104915 Drosophil
c	36	79.8	2.7	421	79	CNS001HF	AL060909 Drosophil
	37	79.8	2.7	1101	79	CNS00IDS	AL075280 Drosophil
c	38	79.6	2.7	833	79	CNS007X3	AL050945 Drosophil
c	39	79.4	2.7	1101	79	CNS003E1	AL064078 Drosophil
	40	79.2	2.7	967	79	CNS00CX5	AL060052 Drosophil
c	41	79.2	2.7	1001	80	CNS0155H	AL105023 Drosophil
	42	79.2	2.7	747	79	AQ331519	AQ331519 nbxb0251I
	43	79	2.7	921	79	CNS0090X	AL052985 Drosophil
	44	78.8	2.7	1101	79	CNS00D0M	AL071284 Drosophil
c	45	78.6	2.7	288	79	CNS007XQ	AL050968 Drosophil

ALIGNMENTS

```
RESULT   1
AA843788
LOCUS       AA843788      467 bp    mRNA            EST       04-JAN-1999
DEFINITION  ak09a04.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            IMAGE:1405422 3', mRNA sequence.
ACCESSION   AA843788
VERSION     AA843788.1  GI:2930239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14  1998 this sequence version replaced gi:1797292
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 779  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1..467
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1405422"
                     /clone_lib="Soares_parathyroid_tumor_NbHPA"
                     /tissue_type="parathyroid tumor"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: parathyroid gland; Vector: pT7T3
                     (Pharmacia) with a modified polylinker; Site_1:
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer
                     [5'-
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTT
                     T-3'], double stranded cDNA was size selected, ligated to
                     Eco RI adapters (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of a modified pT7T3
                     vector (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                     adenomas was kindly provided by Dr. Stephen Marx, National
                     Institute of Diabetes and Digestive and Kidney Diseases,
                     NIH."
BASE COUNT      114 a    111 c     80 g    162 t
ORIGIN
```

```
  Query Match           15.5%;  Score 461;  DB 39;  Length 467;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-76;
  Matches  461;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     598 taacgcattgctctttattaaccc...tgctcaatttatt...tatt 657
           ||||||||||||||||||||||||...|||||||||||||...||||
Db       7 IAACCCATTGCICITTAIIAACCC...ACIICIGICTTA...AIII 66

Qy     658 gagggaagactgttacatgctgtattacaagaaacctgaggctgagatttgaaaaaaaaa 717
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 GAGGGAAGACIGTTACAIGCIGTATTACAAGAAACCIGAGGCIGAGATTTGAAAAAAAAA 126

Qy     718 gaatttaatcaggatgataaaattgaagaagagaaatcagaagttgccaattttcttatg 777
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 GAATTTAATCAGGATGATAAAATTGAAGAAGAGAAATCAGAAGTTGCCAA...TTT...A 186

Qy     778 taagttcccactatctgcaatgttcttttcccagatccttacatgactgattttttt 837
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 IAAGTTCCCACIAICTGCAATGTTCTTTTCCCAGATCCTTACATGACTGATTTTTTT 246

Qy     838 tttattcagatgttaatcagatgtcagcttctcagagaagtgttccttgaaatctctgtaat 897
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 TTTATTCAGATGTTAATCAGATGTCAGCTTCTCAGAGAAGTGTTCCTTGAATTCTCTGTAAT 306

Qy     898 caaagatgcttcccacttcccgactcccaccactagtcactctctgtcccgatattcat 957
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 CAAAGATGCTTCCCACTTCCCGAC...CCA...ACTIAGTCACTCTCTG...A...CAT 366

Qy     958 tttatttccacgtatcactaattcgaattgtattattgaatgtttgtttgtatt 1017
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 TTTATTTTCCACGTATCACTAATT...AATTGTATTATTGAATGTTTGTTTGTATT 426

Qy    1018 ttgtctttctcttaccactcgaatgtatgctccacaggagc 1058
           |||||||||||||||||||||||||||||||||||||||||||
Db     427 TTGTCTTTCTCTTACCACTCGAATGTATGCTCCACAGGAGC 467
```

```
RESULT   2
AQ561028
LOCUS       AQ561028      570 bp    DNA             GSS       17-NOV-1999
DEFINITION  HS_5342_B1_E12_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=918 Col=23 Row=J, genomic survey sequence.
ACCESSION   AQ561028
```

```
VERSION     AQ561028.1  GI:4920696
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3553083.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the Human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.washington.edu/
            Plate: 918  row: J  column: 23
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1..570
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=918 Col 23 Row=J"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      166 a     84 c    122 g    184 t     14 others
ORIGIN
```

```
  Query Match           17.8%;  Score 477.4;  DB 104;  Length 570;
  Best Local Similarity   95.6%;  Pred. No. 4e-75;
  Matches  497;  Conservative    0;  Mismatches   20;  Indels    3;  Gaps    3;

Qy   1300 gaattctctgaagttataatatttgaaagtgaaagcgagagagtctgggacacagctatt 1359
          ||||||||| | |||| |||| || |||  ||||||||||||||||||||||||||||||
Db     45 GNATTCTCTGAAG-TATAATATTNGANAGTGAAAGCGAGAGAGTCTGGGACACAGCTATT 103

Qy   1360 gcttttattcatttttaaaattcacgcaaatcttaaaagcaatacagtgccacaattaagt 1419
          |||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    104 GCTTNTATTCATTTTTAAAATTCACGCAATCTTAAAAGCAATACAGTGCCACAATTAAGT 163

Qt   1420 ggtggctgacttttagagacatgctaaatcctagcagttccactcctagaagagcataat 1479
          |||||||||||||||||||||||||||| | ||||||||||||||||||||||||||||
Db    164 GGTGGCTGACTTTTAGAGACATGCTAAATCCTAGCAGTTCCACTCCTAGAAGAGCATAAT 223

Qy   1480 tagaaaattcaccaaattagctaattatttcaaccaatggtgattgagccgacatgagca 1539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    224 TAGAAAATTCACCAAATTAGCTAATTATTTCAACCAATGGTGATTGAGCCGACATGAGCA 283

Qy   1540 gtgcagttcagactctaccttgacagagcttaaggctttatttcctcaaaaaggaaatgg 1599
          |||||||||||||||||||||||||||| | |||| ||||||||||||||||||||||
Db    284 GTGCAGTTCAGACTCTACCTTGACAGAGCTTAAGGCTNTATTTCCTCAAAAAGGAAATGG 343

Qy   1600 ctaaacgacttgttttatagtctattatgcttaagtagagttttttttttttccataagt 1659
          |||| ||||||||||||||||||||||||||| | ||||| ||||||||||||||||||
Db    344 CTAAACGACTTGTTTTATAGTCTATTATGCTTAAGTAGAG-TTTTTTTTTTTCCATAAGT 402

Qy   1660 tattggggtacaggtggtatttggttacatgactaagttctttagtgtaagtagagattt 1719
```

```
Db    403 TATTGGGGTACAGGTGGTATTT NNNNATGNTAAGT NNNNATNTAANTANATNTN 462

Qy   1720 tatattgcatttttatcatgtg                                    1799

Db    463 TATATTGGCATTTTATGATCTC                                    522

Qy   1800 ttgaaaatacacaaaatagtagagt                                 1859

Db    523 GTG-AATACACAAATAGCACGT                                   571
```

```
RESULT   3
AQ623049
LOCUS       AQ623049      424 bp    DNA                 HTG       PRI 99-SEP-1999
DEFINITION  HS_0342_B2_H01_TFA Homo sapiens DNA clone RPCI-11 Homo sapiens
            genomic clone Plate=918 Col=23 Row=J, HIGH THROUGHPUT SEQUENCE.
ACCESSION   AQ623049
VERSION     AQ623049.1  GI:5085441
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the Human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.washington.edu/
            Plate: 918  row: J  column: 23
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1..424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=918 Col=23 Row=J"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      129 a     70 c     84 g    137 t      4 others
ORIGIN
```

```
  Query Match           12.5%;  Score 335.4;  DB 2;  Length 424;
  Best Local Similarity   99.2%;  Pred. No. 3.3e-44;
  Matches  383;  Conservative    1;  Mismatches    2;  Indels    1;  Gaps    1;

Qy   1335 ataatatttgaaagtgaaagcg                                    1394

Db     40 ATAATATNTGAAAGTGAAAGCG                                    99

Qy   1395 taaaattcacgcaatcttaaaa                                    1454

Db    100 TAAAATTCACGCAATCTTAAAA                                   159
```

```
Qy    1455 agagacatgctaatttctagcagttccactcctagaagagcataattagaaaattcaccaa 1514
           ||||||||||| |||||||||||||||| || |||||||-|||||||||||||||||||||
Db     160 AIAGACAIGCIAAIICIAGCAGIICCACICCIAGAAGAGCAIAATTAGAAAATTCACCAA 219

Qy    1515 attagctaattatttcaaccaatagtcattgagccgacatgagcagtgcagttcagactc 1574
           ||||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     220 AIIAGCIAAIIAIIICAACCAAIGGICAIIGAGCCGACATGAGCAGIGCAGTTCAGACTC 279

Qy    1575 tacottgacagagcttaaggcttttatttcctcaaaaaggaaatggctaaacgacttgtttt 1634
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 IACTTIGACAGAGCTIAAGGCIIIAIIICCTCAAAAAGGAAATGGCTAAACGACTTGTTT 339

Qy    1635 tatagtctattatgcttaagtagagtttttttttttttccataagttattggggtacaggt 1694
           |||||||||||||||||||||||| |||||.|||||||||||||||||||||||||||||
Db     340 IAIAGICIAIIAIGCIIAAGIAGAG-IIIIII:IIIICCATAAGIIAIIGGGGIACAGGT 398

Qy    1695 ggtatttggttacatgagtaagttct 1720
           |||||||||||||||||||||||||
Db     399 GGIAIIIGGIIACAIGAGIAAGIICI 424


RESULT   4
D61567
LOCUS       D61567        302 bp    mRNA            EST       14-DEC-1995
DEFINITION  HUM418G12B Clontech human fetal brain polyA+ mRNA (#6535) Homo
            sapiens cDNA clone GEN-418G12 5', mRNA sequence.
ACCESSION   D61567
VERSION     D61567.1  GI:970692
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 302)
  AUTHORS   Fujiwara,T., Hirano,H., Kategiri,I., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Fujiwara et al. (1995)
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785681.
            Contact: Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 Kagasuno, Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
            Tel: 0886-55-2888
            Fax: 0886-37-1035
            Insert Length: 1881  Std Error: 0.00
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1..302
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GEN-418G12"
                     /clone_lib="Clontech human fetal brain polyA+ mRNA
                     (#6535)"
BASE COUNT       79 a     78 c     65 g     79 t      1 others
ORIGIN


  Query Match           10.2%;  Score 311.2;  DB 24;  Length 302;
  Best Local Similarity   99.7%;  Pred. No. 3.2e-46;
  Matches  301;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 tgtatttccagttaactgtcagaagagaaat:ctgaagagctgtagtcgtgcatcctttc 85
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1 IGIAIIICCAGIIAACIGCICAGAAGAGAGA:GITGAAGAGCIGIAGICGIGCATCCIIC 60

Qy      86 tcaccctccgttagaaagcctcctctcatcctcagaagactactgtcagaggatgtagga 145
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IIAICCIICCIIIAGAAAGCCICCICIICAICCICAGAAGACIICCIICAGAGGAIGIAGGA 120

Qy     146 atggaatcccctttgaagaggggtgtcgaatccccagtgctgcagacatgaggcctgaa 205
```

```
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGGACATCCCCTTTGAAGAGGGCGTCGAGTCCCAGTGCTGCAGACATGAGGCCTGAA 180

Qy     206 cctcctaattctctggatcttaatagactcatcctcgggaaatcaagctgcctgcctcc 265
           |||||||||||||||||:|||||  ||||||||||||||||||||||||||||||||||||
Db     181 CCTCCTAATTCTCTGGGVICTTAAIGACACICAICCTCGGAGAAICAAGCTIACTGCTTA 240

Qy     266 aatatcaatctttctctggaccaaagtgaaggatctattctctctgatgatgatgtttcc 325
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATGATGTTTCC 300

Qy     326 ag 327
           ||
Db     301 AG 302


RESULT   5
AA504063
LOCUS       AA504063      275 bp    mRNA            EST       22-AUG-1997
DEFINITION  nh40g09.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954880, mRNA
            sequence.
ACCESSION   AA504063
VERSION     AA504063.1  GI:2239030
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811027.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Podrag F. Tlaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 404  Std Error: 0.00
            Seq primer: -40m13 fwd  BI from Amersham
            High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1..275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:954880"
                     /clone_lib="NCI_CGAP_Pr5"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from normal prostatic
                     epithelial cells, cDNA made by oligo-dT primime.
                     Non-directionally cloned. Size-selected >400 bp. To get
                     average insert size 600 bp. "
BASE COUNT       80 a     66 c     62 g     67 t
ORIGIN


  Query Match           8.7%;  Score 256.4;  DB 34;  Length 275;
  Best Local Similarity   99.6%;  Pred. No. 2.6e-38;
  Matches  259;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     128 gagtactgtcagaggatgtaggaatccc:t:c:ttgaaaaggcctgaa 192
           |||||||||||||||||||||||||||| | |||||||||||||||||
Db      16 GACTACTGTCAGAGGATGTAGGAATCCAGTCCTTTGAAGAGGCCTGAA 75
```

```
Qy     183 gtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcctc 242
           ||||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||||
Db      76 GTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTC 135

Qy     243 ggaggaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatcta 302
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 GGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 195

Qy     303 ttctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttg 362
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     196 TTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTG 255

Qy     363 ataccccccgatgaagcagat 382
           ||| ||||||||||||||||
Db     256 ATAGCCCCCGATGAAGCAGAT 275


RESULT   6
AQ452851/c
LOCUS       AQ452851      429 bp    DNA             GSS       23-APR-1999
DEFINITION  HS_5207_A2_F10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=783 Col=20 Row=K, genomic survey sequence.
ACCESSION   AQ452851
VERSION     AQ452851.1  GI:4625621
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3897
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 783  row: K  column: 20
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1..429
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=783 Col=20 Row=K"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      137 a     85 c     90 g    115 t      2 others
ORIGIN


  Query Match           6.2%;  Score 194.8;  DB 103;  Length 429;
  Best Local Similarity   87.0%;  Pred. No. 1e-24;
  Matches  214;  Conservative    0;  Mismatches   29;  Indels    3,  Gaps    1.

Qy    1072 ttgatcctgcgtctccaataaagtgggaaggaaggaaggaaggatggtttaggggagtgaga 1131
           || || |||||||| | ||   | | | | | ||||| ||||||||||||||||||||||
```

```
Db     245 TTGGTCNTGCGTCTCAATAAA---G GGAAGGAAGGAAGGAAGGATGAGA 189

Qy    1132 gaaatgataaaaaaggaggataatttaaaagtttcagatcatcaaaaaaaaaaaaaaaaa 1191
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     188 GAATTGATAAAAAGGAGGATATTTAAAGTTTCAGATCATCAAAAAAAAAAAAAAAAA 129

Qy    1192 tatgaacttcagtctttctttttttttttttttttttttttttttttttttttttttt 1251
           ||||| ||||||||| | | | | | | | | | | | | | | | | | | | | | |
Db     128 TATGAGCTTCAGTCTGTCATAAGTTCTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTT 69

Qy    1252 atgtacaaatacaatcatttttttttttaaatgtgtattatgttaattttgtgtttcagtttctag 1311
           |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      68 ATGTACNAACACAATCATTTTTTTTTTAAATGTGTATTATGTTAGTTTTGTGTTTCAGTTTCTAG 9

Qy    1312 ccaggg 1317
           |||||| |
Db       8 CCAGGG 3


RESULT   7
AA296475
LOCUS       AA296475      236 bp    mRNA            EST       11-APR-1997
DEFINITION  EST10971 Umbilical vein endothelial cells II Homo sapiens cDNA 5'
            end, mRNA sequence.
ACCESSION   AA296475
VERSION     AA296475.1  GI:1948819
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Mac-Wall,T.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.F., Liu,L.-I., Markham,P.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,T.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Sadow,P.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,J., Depinet,M.A., Gelmann,E.A., Phillips,B.C.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,T., Hastings,C.A.,
            He,W.W., Hu,J.S., Groene,E.M., Gruber,J., Wilson,F., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Kung,J.T., Lillie,A., Agassandian,K.S.,
            Raymond,L., Wei,G.F., Wang,J., Xu,C., Yu,G.L., Fuhrer,S.M.,
            Dillion,P.J., Fannon,M.S., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl.), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Apr 14, 1999 this sequence version replaced gi:1948819.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please refer to the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..236
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Umbilical vein endothelial cells II"
                     /cell_type="endothelial cell"
                     /dev_stage="fetus"
```

```
                     /note="Organ: umbilical vein; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       66 a     54 c     57 g     55 t      4 others
ORIGIN


  Query Match             6.0%;  Score 179.4;  DB 31;  Length 236;
  Best Local Similarity   97.8%;  Pred. No. 1.1e-23;
  Matches  180;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy     127 gaatactatcagaggtatatggaatggacatc.cctttgaagaggcgtgctgagtccca 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      58 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCA 117

Qy     183 gtgctgtgagacatgaggcctgaacctcttaattctctggatcttaatgacactcatcctc 242
           |||||||||||||||||||||||||| ||| || ||| |||||||||||||||||||||
Db     118 GTGCTGCAGACATGAGGCCTGAACCTCNTAATTCTNTGGTTCTTAATGACACTCATCCTC 177

Qy     243 ggaggaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatcta 302
           |||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db     178 GGAGAATCAAGCTCACAGCCCCAAAATNTCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 237

Qy     303 tttt 306
           ||||
Db     238 TTCT 241


RESULT   8
AQ533091/c
LOCUS       AQ533091       565 bp    DNA             GSS       18-MAY-1999
DEFINITION  RPCI-11-351B1 TV RPCI-11 Homo sapiens genomic clone RPCI-11-351B1,
            genomic survey sequence.
ACCESSION   AQ533091
VERSION     AQ533091.1  GI:4844781
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 868 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..565
                     /organism="Homo sapiens"
                     /db_xref="GDB:7634424"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-351B1"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      142 a    150 c    114 g    159 t
ORIGIN
```

```
  Query Match             3.6%;  Score 106.8;  DB 104;  Length 565;
  Best Local Similarity   90.5%;  Pred. No. 2.6e-10;
  Matches  114;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    0;

Qy      18 ggcgaatctgtatttccagttaagt.tctcagaagagagatgctgaagaggctgtgatgcta 77
           ||||||||||||||||||||||||||.|||||||||||||||||||||||||||||||||
Db     230 GGCGAATCTGTATTTCCAGTTAAGTTTTTCAGAAGAGAGATGCTGAAGAGGCTGTGATGCTA 171

Qy      78 catcctttctcacctccgttagaaa.ttcct.tcatcttcaggggctg.tatgaaga 137
           |||||||||||||||||||||||||||||||||||||||||||||||| |||| ..||
Db     170 CATCCTTCTCACCCTCCGTTAGAAAGCTTCCTCTCATCCTCAGAAGGTAGGTTTTTTT 111

Qy     138 atgtag 143
           ||||||
Db     110 GTGTAG 105


RESULT   9
CNS00EVL/c
LOCUS       CNS00EVL     1101 bp    DNA             GSS       4-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resources Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_rpci98.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN


  Query Match             3.1%;  Score 92.8;  DB 79;  Length 1103;
  Best Local Similarity   26.2%;  Pred. No. 9.3e-08;
  Matches  160;  Conservative   85;  Mismatches  197;  Indels    0;  Gaps    0;

Qy    2312 atagctatatttttgtaaaacatgaaa aggctaagtga tatatttttttttttt  2371
           ||  | |||||| || | |||||||||| ||| |||||||||||||||||||||
Db    1096 ATWTWTTTWTTTTTWWTATWTTAWATATATAATTAATTWWTWTATATWTAWGWTNTWATA 1037
```

```
Qy   2372 taaatatgtttgagttctgaagtatgaaataaatgcttgaaagcagcctgtttcttaggt 2431
Db   1036 TATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTAT 977

Qy   2432 ttctcacggcgaccatttttgaaccaggaatgtacattctaaacaaaaatagggtttttaag 2491
Db    976 AWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTW 917

Qy   2492 gtatagtttttcaaaaacatttgggccataatttatttttcttttaatgaatacttttttca 2551
Db    916 TTTWTATWIWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAW 857

Qy   2552 tttatatttatatttttcgtgtatatatttttcattatatacgtttcattatatattatata 2611
Db    856 ATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTW 797

Qy   2612 tttgtactatatatttttcatttatatagttttcattgtatattatatacctcattatat 2671
Db    796 TATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTT 737

Qy   2672 tatataattcattatatagttttcatttaatttatttcatttttataactagaacaatgaa 2731
Db    736 TATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTAT 677

Qy   2732 gcacagagaagtttaagtaattt 2753
Db    676 TTAATTAWWWAATWAWTTATTT 655
```

RESULT 10
CNS00Z6Z/c
LOCUS CNS00Z6Z 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN01A10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097301
VERSION AL097301.1 GI:5609912
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN01A10"
/note="end : T7"
BASE COUNT 432 a 108 c 89 g 253 t 219 others
ORIGIN

Query Match 3.1%; Score 91.8; DB 79; Length 1101;
Best Local Similarity 33.8%; Pred. No. 1.4e-07;
Matches 158; Conservative 104; Mismatches 207; Indels 0; Gaps 0;

```
Qy   2251 aatttttaaaatgaaagatatatatttttttgtttgaaatttttaaaaaaaaaaaaaaaa 2309
Db    999 AAAAAWWTWAWATTTWTAWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTW 940

Qy   2310 gaatagctatatttttgtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaat 2369
Db    939 NDMBDTGTTTTMHTKTKKTRMWSCNSSSTSTTTVCHKAAAAAAAAAAAAAAAAAAAAAA 880

Qy   2370 tataaatatattttgagttctgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2429
Db    879 DAATWATTTWTRWWWWTWAWWWTWAWTWTWTWTWTWWWWWWWWWWWWWWWWWWWWWWWW 820

Qy   2430 gtttctcaaggcgaccatttttgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2489
Db    819 GTWTTTPMVAAAAAAATAAAATWWAAAAWVTTTWTMCWTWTWWWWWWWWWWWWWWWWWTA 760

Qy   2490 aggtatagttttttcaaaaacatttgggccataatttatttttcttttaatgaatacttttt 2549
Db    759 GRKTAATAHDAAMTAWRAWWTWWTTKVTKMTWWATTTTTTVTWTWTWTTTTTTTTTTTTA 700

Qy   2550 catttatattatatttttcgtgtatatatttt atgtatata gttt tttcatttatttata 2609
Db    699 MTMKKTAAAAMAAKTKTWTKMTKTKWWWTTKRMCAAWAWTTTTTTTTTTTTTTTTTTTA 640

Qy   2610 tatttgtactatatatttttcatttatagttttcatttgtatattatatacctcattatat 2669
Db    639 ATATAAWTATAYYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA 580

Qy   2670 attatataattcattatatagttttcatttaatttatttcatttttat 2716
Db    579 TTTTTTTTTTTTTTTYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 533
```

RESULT 11
CNS00LT2/c
LOCUS CNS00LT2 1101 bp DNA GSS 14-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL078714
VERSION AL078714.1 GI:5102074
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The RPCI Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2, cn bw sp. The sequences sent to the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resources Center can be found at http://bacpac.med.buffalo.edu/projects/drosphil.htm.
FEATURES Location/Qualifiers
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

```
                     /clone="BACR48P19"
                     /note="end : TET3"
BASE COUNT      469 a      6 c     69 g    151 t    406 others
ORIGIN


  Query Match             3.1%;  Score 91;  DB 79;  Length 1101;
  Best Local Similarity   42.5%;  Pred. No. 2e-07;
  Matches  184;  Conservative   49;  Mismatches  200;  Indels    0;  Gaps    0;

Qy   2284 aaactttgaaaacacaaagtgcctctgaata:tatatttttgtaaacatgaggacatgga 2343
          ::::::|  ::::::|||  :   : : :  :   | |||||  ||||  |   |  : :|  |
Db    715 HMWMHTTTTMMMMMAAAWWTTTWAHAAAWTTTITTATTTAATAAAAAAAAAAMAMTAAA 656

Qy   2344 taagtgactgtataagagggttggattataaatatgtttgagttctgaagtatgaaataa 2403
           ||   : :  : || |     |  |  | :|| |    : | | :: :: : : ::  |:
Db    655 AAAAAAATWTAWAAAWAAAAATTWAAAAAWAAAAAWAAWAAATAHWTWWMWTWAMMTAAW 596

Qy   2404 atgcttgaaagcagcctgtttcttaggtttcaggcgacccattttgaaccaggaatgta 2463
          :     |   :  : :|: | :   |: :: :  :   :   || |  | |
Db    595 AMAAAAWATTAAAAAAWWTTWTTAMTTAMTMMWHAMATMWAAWMAAAAAAAATATTAAT 536

Qy   2464 cattctaaacaaaatagggttttaaggtatagttttcaaaaacatttgggccataat 2523
           | :  : | |||::    :|||:|| :| | |  :  ||||| | ||      |  |
Db    535 WAAWAAWTAAAAAHWWTTWMHHTTHAAAAWAAAATMMWAAAAAAAWAATTTTTNTTTT 476

Qy   2524 ttatttttttttaatgaatactttttcatttcattatatttttcgtgtatatattttca 2583
          || :| | |||,  |   |  ||||   ||  | || | ||||  | | | | ||||
Db    475 TINTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 416

Qy   2584 ttatatagtttcattatatattatatatttcactatatatttttcatttatatagttt 2643
          || | |  |||  || | | || | | ||| |  | | | |||||  ||| | |  |||
Db    415 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 356

Qy   2644 tcattgtatatttatatacctcattatatttatcaattcattatatagttttcatttaatt 2703
            || | | || | |   | || | |  |    ||  || | |  |||| |||  ||
Db    355 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 296

Qy   2704 tatttcatttttat 2716
          | |||  |||| |
Db    295 TTTTTTTTTTTTT 283


RESULT  12
CNS0039G/C
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence IET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2, cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of this library
            and how to order individual BAC clones, the entire library or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a      64 c    131 g    202 t    503 others
ORIGIN


  Query Match             3.0%;  Score 88;  DB 79;  Length 1101;
  Best Local Similarity   18.5%;  Pred. No. 7.2e-07;
  Matches  126;  Conservative  280;  Mismatches  276;  Indels    :  Gaps    0;

Qy   2005 tttttgtcaggagagctttgtcaacacaatctgaacctttttatttttaaaaatatttgtt 2064
          ::|::::: |:::::  : ::: : :   ::   ::  ::::::|:|:::  ::: :|
Db   1077 WWTWWKDPADDPPWAGDADPWAWDDCACTTWTATWWWWWWWWATWDTWWPKWWWWATAAK 1018

Qy   2065 actgtcgagcaaagcagtggtggctgggcaaagctgctaatttaacactaaaattatttat 2124
           |:  :: |::  :  :: |  :::: |  :: :|  :: :  :: : :: ::  :  ::
Db   1017 TDTAWTWWPTAWPADWAGPDPGAGPDPDAATPADCACPPPDPPPPKPWWPPKPDPWPPPWK 958

Qy   2125 atggctcttttaatgtagttaacatggctatgttgaaagaaccaattttaatttccaaaaat 2184
          :: :  :  :  :: :::: ::: :: ::: |::::  ::::: :|:  : :
Db    957 GGKKKKAAKAAKWATKWWDDWDWDKDWWDGAPDPKATTTDGADTPTTTDKPDADDDVP 898

Qy   2185 cctgttttaggtcttaaaatgcaaaggacagaattgtaaactaaatttaaattttttaaatat 2244
          :::::::: : :: |   :|  | |  ::: : :|  :: ::|  :::  :::: :: :
Db    897 GTKDDDDKDKWDDWDKAKGTWDDATWAWAATTDWWWGWADADWWTWPAAAPTWPPPAWP 838

Qy   2245 accttaattttcaaatgaaagatatatagtcaaaggttgaaaatttaaaaaaacaaaatat 2304
          | ..... .  :.| ...:: :    :  :: ::: :: : : :  :::: : : :
Db    837 AWAWKWDDAWAWGAPTADPPDWWDPDPDDDADPKPPDPPPAPPPPDAAPPPPPAATWTWW 778

Qy   2305 cctctgaatagctatattttgtaaacatgaggacatggataagtgtgctatatgggtttt 2364
          | |:: ::  : :|:|:   ::: :::::: ::  :  :: : :  :  :::::::::
Db    777 TTTTIRDTDDWKWKTDTWTPWAADPTWPPDDDDDPDPAGTAGPKWPPTWKPPPWKPRITPW 718

Qy   2365 tggattataaatatgtttgagttctaaagtataaaataaatgatttaaaaaajcctttt 2424
          :     :: ::: : :  :  :  :  .....| :: :: :::: :  : : :
Db    717 DDADADDTAPDDPPPPGDDGADADKPPTTPPPPPPTPATWTPTTAWWPPAAWWTTTDTW 658

Qy   2425 cttaggtttctcacgcgaccat'''aaaaagaaaatgtaaatttaaaaaaaaaaaaaaa 2484
          :::::  :    :  ::      :: :: ::: :: :: :::: :: :  :  :
Db    657 TDDWDKPDPPPKGAPPPPTTAPAAWDWWTWPAWDWAKWDWKTPADPWDPWAAPTWTPAA 598

Qy   2485 ttttaaggtatagttttttcaaaaaaatttgggcatgaatttatttttttttaaatgtaa 2544
          : .... :  ...  :  :|||: : :| :|   || :::::  :  :::::  :
Db    597 KADPDWAKAPAWPAPPDPAPAAPADPPWTTKGPTTTATWTTWAAPAAWKWWAWATTTAT 538

Qy   2545 tttttcatttatattatattttttgtgtatatttttcatttatatatatttttaaaaaa 2604
          :|||:  |:| | || | ||:|'   :: |:|| :: |:| |:  :  :
Db    537 WTTTWTTTWTTTTTTTTTTWTTTAAWWPAWWTATWAAWTAAAWAAAAAAAAAAAAAAA 478

Qy   2605 ttatatatttgtactatatatt'''''''''''''''''''''''''''''''''''''' 2664
          || | | :  | .:||. .| |   '':  :  :|||  :  :  :
Db    477 TTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTTWTTT 418

Qy   2665 attatattatataattcattat 2686
          :| | |  | .....|  |||
Db    417 WTAAAAAAAAWWWWWTWTATAT 395


RESULT  13
CNS0021J/C
LOCUS       CNS0021J     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence IET3 end of BAC #
```

```
            BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061936
VERSION     AL061936.1  GI:4940214
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR05N11"
                     /note="end : TET3"
BASE COUNT      631 a      7 c     28 g    289 t    146 others
ORIGIN
```

```
  Query Match             3.0%;  Score 87.6;  DB 79;  Length 1101;
  Best Local Similarity   44.2%;  Pred. No. 8.5e-07;
  Matches  236;  Conservative   33;  Mismatches  261;  Indels    4;  Gaps    1;

Qy    2197 tgttttaggtcttaaatgcaaaggacagaattgtaaactaagttgaatttttaaatctac 2246
           | ||||   | || |||  |||  |   |     |||      |: |||:|| : || ||
Db     818 TTTTTTTTTTTTTTAATTAAAAAAAATTAAATTAAAAAATTTTWTAATWTTAWTATWTAA 759

Qy    2247 cttaatttttcaaatgaaagatatatagtcagaggttgaaactttgaaaacacaaagtgcc 2306
             | || : |: :  :   || | :     |  ||  || :  ||||   |||
Db     758 AAAATTTAWWAWTWATWTTTTAAAAWTAATAAAWTTTTAAAWAAAAAAAWTAAAAAAWTT 699

Qy    2307 tctgaatagctatatttttgt----aaacatgaggacatggataagtgactgtataagagg 2362
           | | ||    :|| |: | :    |:| :: :     |  |:  | || |  |:|| |
Db     698 TTTAAAATTTWATTTWATWWATAAAWAAWTTWTATWTTTTTTWTATTAATTAAWAAAAAA 639

Qy    2363 gttggattataaatatgtttgagttctgaagtatgaaataaatgcttgaaagcagcctgt 2422
           .  .| |||| .| ||| .. | .| :|  |:  :|  ||:  :|  | |  | |  | :
Db     638 AAWAWTWTTTAAAAWTTTTTTTTWWTTAWATTAAAAWTAAAWAWTWTTTTATTAAATTTW 579

Qy    2423 ttcttaggttttctcagcgacctattttgaaacaggaatgtacattctaaacaaaataggg 2482
           |: | |   ||   |    : |     ||  :  ||  :  |||  :|| |||:||
Db     578 TWTTAAAAATTAAAATTAAWAAAAAAAAAAAAAAAAAAAAWTAATTAAWAAAAAAWTAAAT 519

Qy    2483 gttttaaggtatagttttttcaaaaacatttgggccataatttatttttcttttaatgaat 2542
           | ||    |   ||||      |||       |  ||| |||| || |  |   |
Db     518 TATATTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNTTTTTTT 459

Qy    2543 actttttcatttatattatattttcgtgtatatattttcattatatacgtttcattata 2602
           |||||  ||| | ||  |||||  '     | ||||  || | |   |||  || |
Db     458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 399

Qy    2603 tattatatatttgtattatattttttcattttatatttttcttttatatttatttatttat 2662
           | |' | | |||| | ' ' '
Db     398 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 339

Qy    2663 tcattatattatataattcattatatagttttcatttgattttgttgattttat 2716
           | ' | | |' ' ' '  '  '  '  '  '  '  '  '  '  '  '  '
Db     338 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 280
```

```
RESULT  14
CNS010MP/c
LOCUS       CNS010MP      734 bp    DNA              GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence, End of BAC
            BACN04E20 of DrosBAC library, from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billault at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04E20"
                     /note="end : T7"
BASE COUNT      238 a     62 c     [illegible] g    221 t    [illegible] others
ORIGIN
```

```
  Query Match             2.5%;  Score [illegible];  DB [illegible];  Length [illegible];
  Best Local Similarity   89.2%;  Pred. No. 2.5e-[illegible];
  Matches  198;  Conservative   [illegible];  Mismatches   22;  Indels    [illegible];  Gaps    2;

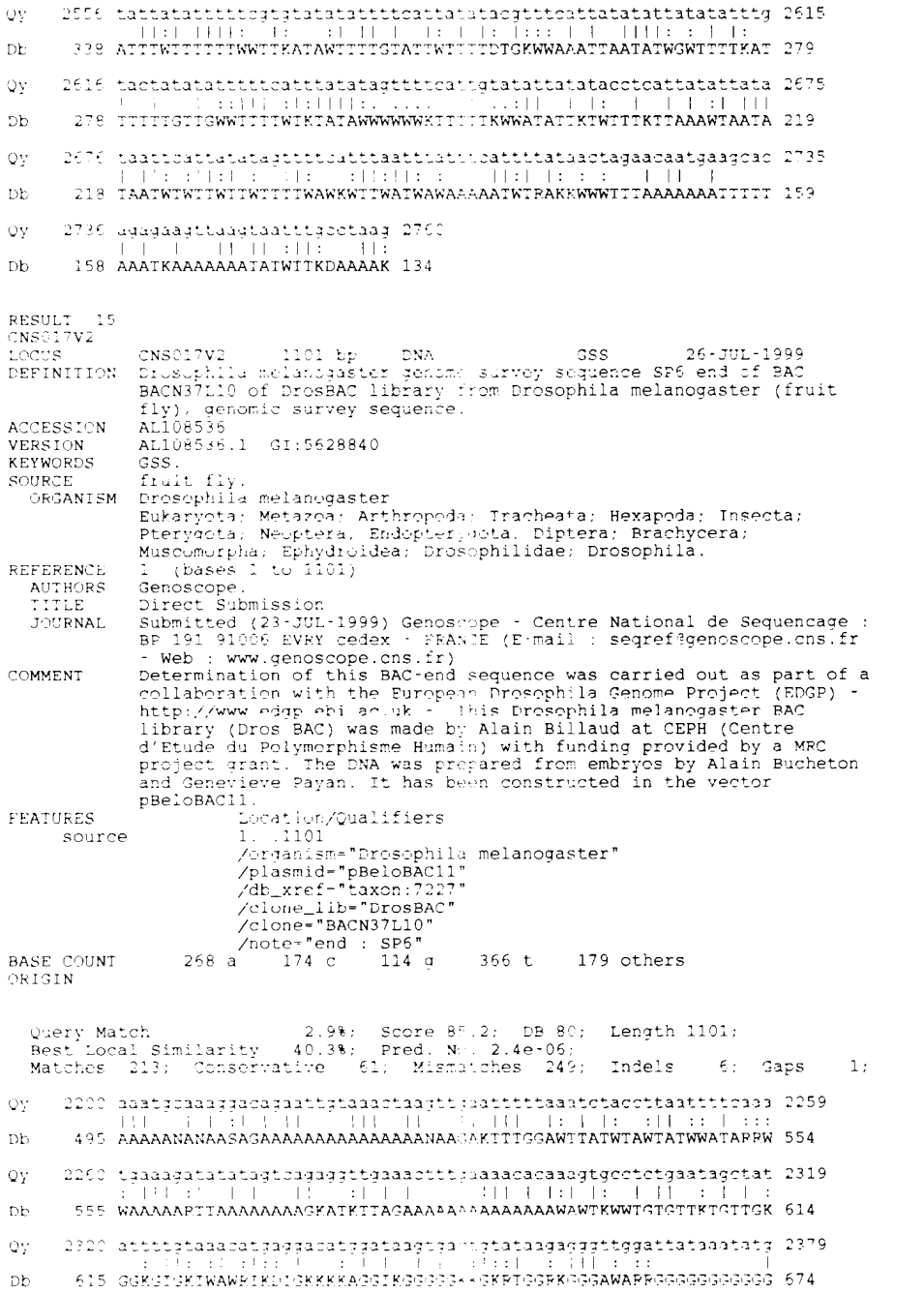
Qy    2269 aagatatatagtcagaggttg[illegible] 2322
           || | | |  |  |  | | :  :  :  :  :  :  :  :  :  :  :  :  :
Db     638 AAMMAATGAATTTKAATCTWAA[illegible] 579

Qy    2323 ttgtaaacatgaggacatggat[illegible] 2379
           |:|   | |  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     578 GTDTTTSATTAAAAAATWDGDGT[illegible] 519

Qy    2380 tttgagtttgtgaagtttg[illegible] 2439
           |::  :| | |  |  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     519 GTKKGTKTTTGATAAAATAKTTTTTTTTTTSTGGGAATTTRTTTTNTTTAATTTTAAAAA 459

Qy    2440 cgaccatttgaaacaggaa[illegible] 2495
           |  : | ||||  | ||  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     458 AAATGDGTGTGAAGTATTAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 399

Qy    2496 agttttttaaaaaacatttgggc[illegible] 2555
           :||''  :: |  :|: | |  :  :  :  :  :  :  :  :  :  :  :  :
Db     398 TCTTATAAWWTATKWATWTTGTAAAAAWTWTTTTATTKTTTTTTTTTTTTTTTTTTTTT 339
```



```
RESULT  15
CNS017V2
LOCUS       CNS017V2      1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37L10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108536
VERSION     AL108536.1  GI:5628840
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37L10"
                     /note="end : SP6"
BASE COUNT      268 a    174 c    114 g    366 t    179 others
ORIGIN


  Query Match             2.9%;  Score 85.2;  DB 80;  Length 1101;
  Best Local Similarity   40.3%;  Pred. No. 2.4e-06;
  Matches  213;  Conservative   61;  Mismatches  249;  Indels    6;  Gaps    1;
```



```
Search completed: February  7, 2000, 19:33:28
Job time: 6718 sec
```

```
OM of: US-09-174-937-8 to: GenEmbl:*   out_format : pfs

Date: Feb 9, 2000  4:06 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=  /SPTI_gcg_US-09-174-937/raw_  /122[1]_1451/8_13281/app_query.fasta 1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -XGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block.
Query: US-09-174-937-8
Query length: 126
Database: GenEmbl:*
Database sequences: 821193
Database length: 1778182014
Search time (sec): 1185.490000



score_list:
Sequence              Strd Orig    ZScore    DScore Len  ! Documentation ..
gb_pr1:AB002365         -  484.00  962.93  4.8e-40   5654  ! AB002365 Human mRNA for KIAA0
gb_htg6:AC009556        -  477.00  934.60  1.1e-39  168798  ! AC009556 Homo sapiens chromo
gb_htg3:AC011489        -  261.00  425.96  1.1e-15  179301  ! AC011489 Homo sapiens chromo
gb_pat:AR028453         -  147.00  256.32  3.3e-06    1985  ! AR028453 Sequence 12 from pate
gb_pr2:HSU15172         -  147.00  251.99  5.0e-06    2982  ! U15172 Homo sapiens BCL2/adeno
gb_ro:AF035207          -  135.00  220.44  9.2e-05    2213  ! AF035207 Mus musculus Nip21 mR
gb_htg6:AC009556        -  118.00  201.70  1.4e-01  168798  ! AC009556 Homo sapiens chromo
gb_htg5:AC009481        -   92.50  115.88  5.4e+00  192354  ! AC009481 Homo sapiens clone
gb_pr2:HSU35709         -   89.50  111.27  9.6e+00  171269  ! U35709 Human chromosome 17p1
gb_pr3:HS357116         -   89.50  113.25  8.5e+00  134506  ! AL021921 Homo sapiens DNA se
gb_ba1:AB017508         -   88.00  123.10  1.7e+00   30050  ! AB017508 Bacillus halodurans
gb_ro:AF090436          -   84.50  137.51  12.31     2256  ! AF090436 Mus musculus dachshun
gb_ro:AF090437          -   84.50  139.10  11.42     2100  ! AF090437 Mus musculus dachshun
gb_ro:AF102547          -   84.50  137.51  12.31     2256  ! AF102547 Mus musculus dachshun
gb_ro:MMU05669          -   84.50  136.79  12.52     2467  ! AF005669 Mus musculus mRNA for
gb_ro:MMTRPTP           -   83.00  127.19  46.29     5691  ! Z22794 M.musculus trp protein 1
gb_htg5:AC014844        -   82.50  127.79  [illegible]  54445  ! AC014844 Drosophila melanogas
gb_pr4:AC006474         -   81.50  103.93  913.93    69718  ! AC006474 Homo sapiens clone D
gb_in1:S66759           -   81.00  134.14  18.98      1353  ! S66759 TFIIA-L-transcription f
gb_ba1:SC10H5           -   80.00  122.95  [illegible]  4870  ! AL031232 Streptomyces coelicol
gb_pr2:AP000035         -   80.00   98.22  1.0e-02  100000  ! AP000035 Homo sapiens genomic
gb_pr2:AP000101         -   80.00   98.22  1.0e-02  100000  ! AP000101 Homo sapiens genomic
gb_pr2:AP000177         -   80.00   98.22  1.0e-02  100000  ! AP000177 Homo sapiens genomic
gb_pr2:AP000264         -   80.00  102.34  1.1e-03   60455  ! AP000264 Homo sapiens genomic
gb_pr2:HSU20980         -   80.00  129.46  31.57      2197  ! U20980 Human chromatin assembl
gb_pl1:ATF20P19         -   79.50   96.30  1.0e-02  134732  ! AL049483 Arabidopsis thaliana
gb_pl1:ATU12959         -   79.50  [illegible]  [illegible]  1169  ! U12959 Arabidopsis thaliana C
gb_pr3:HSU47741         -   79.50  117.29  264.74     8694  ! U47741 Human CREB-binding prot
gb_in1:CEF54E4          -   79.50  104.18  844.95    43150  ! Z79639 Caenorhabditis elegans
gb_pr4:HSU85962         -   79.50  117.82  [illegible]  5147  ! U85962 Homo sapiens IREB-bindi
gb_htg6:AF127019        -   79.50   92.38  4.0e-02  192405  ! AF127019 Homo sapiens chromo
gb_htg1:HsBA388K1       -   79.00   91.91  4.1e-03  174795  ! AL121584 Homo sapiens chromo
gb_htg2:AC007449        -   79.00   91.96  4.3e-03  173727  ! AC007449 Homo sapiens clone
gb_in2:AC004423         -   79.00   98.30  1.9e-03   79076  ! AC004423 Drosophila melanogas
gb_in2:AF132884         -   79.00  124.93  61.90      3053  ! AF132884 Drosophila melanogast
gb_htg5:AC016128        -   78.50   98.57  1.8e-03   69431  ! AC016128 Homo sapiens chromos
gb_htg5:AC016128        -   78.50   98.57  1.8e-03   69431  ! AC016128 Homo sapiens chromos
gb_ro:MMU85713          -   78.00  121.06  102.55     3915  ! U85713 Mus musculus phospholip
gb_in2:AF203842         -   78.00  117.33  109.90     6178  ! AF203842 Drosophila melanogast
gb_pl1:SPCC1006         -   77.00  101.99  1.0e-02   32000  ! Z98951 S.pombe chromosome I c
gb_pl1:SPCC1393         -   77.00  102.32  1.1e-03   30309  ! AL031546 S.pombe chromosome I
gb_ro:AB019120          -   77.00  117.30  [illegible]  4351  ! AB019120 Rattus norvegicus mRN

gb_htg1:CEV43H4         -   77.00  [illegible]  9.4e+00  [illegible]  ! AL[illegible] Caenorhabditis
gb_htg2:AL133293        -   77.00   94.14  3.2e-03   [illegible]  ! AL[illegible] Homo sapiens ch
gb_in1:CEE47E8          -   77.00   93.[illegible]  1.7e-03  [illegible]  ! [illegible] Caenorhabditis el

seq_name: gb_pr1:AB002365

seq_documentation_block:
LOCUS       AB002365     5654 bp    mRNA            [illegible]  [illegible]-1999
DEFINITION  Human mRNA for KIAA[illegible] gene, partial cds
ACCESSION   AB002365
VERSION     AB002365.1  GI:[illegible]
KEYWORDS    KIAA[illegible].
SOURCE      Homo sapiens male brain cDNA to mRNA, [illegible] clone:HHC141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Nakajima,D., Kikuno,R.,
            Miyajima,N., Kotani,H., Nakayama,M. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1998) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Kisarazu, 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:[illegible]@kazusa.or.jp,
            URL:http://www.kazusa.or.jp [illegible])
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="[illegible]"
                     /clone="HHC141"
                     /clone_lib="[illegible]"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA[illegible]"
     CDS             <1..2151
                     /gene="KIAA[illegible]"
                     /codon_start=1
                     /product="[illegible]"
                     /db_xref="G:[illegible]"
                     /translation="[illegible]HSPSDASG
                     [illegible]SQNIDEDLL
                     [illegible]SPEDEELGC
                     [illegible]SPEDEELGC
                     [illegible]ASNILSPSN
                     [illegible]EERSTEPN
                     [illegible]LASFPQTC
                     [illegible]ASPDAVTH
                     [illegible]ARDSGPGWS
                     [illegible]SPDEIDINV
                     [illegible]LYVISTLE
                     [illegible]"
BASE COUNT      [illegible]
ORIGIN


alignment_scores:
            Quality:  484.00            Length:    [illegible]
              Ratio:  5.204             Gaps:      [illegible]
 Percent Similarity: 100.000   Percent Identity: [illegible]

alignment_block:
 US-09-174-937-8 x AB002365   [illegible]
```

```
Align seq 1/1  to  AP000365  from  1  to  5654

      24 LeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValIle 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1471 CTACTGTCAGAGGATGTAGGAATGGACATCCCTTTGAAGAGGGCGTGCT 1520

      40 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 57
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1521 GAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCCCTGGATC 1570

      57 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 73
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1571 TTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATATCAAT 1620

      74 LeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAs 90
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1621 CTTTCTCTGGACCAAAGTGAAGGAATCTATTCTTCTGATGATAACTTGGA 1670

      90 pSerProAspGluIleAspIleAsnValAspGluLeuAspThrProAspG 107
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1671 CAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCGGATG 1720

     107 luAlaAspSerPheGluTyrThrGlyHis 116
         |||||||||||||||||||||||||||||
    1721 AAGCAGATTCTTTTGAGTACACTGGCCAT 1749

seq_name: db_htg6.AC009556

seq_documentation_block.
LOCUS       AC009556   169798 bp    DNA              HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479108
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colanuelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeAngelano,K., Dempsey,E., Devon,K., Dewar,K.,
            Donelan,J., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pillara,V., Riley,E., Roberts,D., Roy,A., Severy,F.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13, M77815; 1.0% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator-Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
*          1     1016: contig of 1016 bp in length
*                      gap of unknown length
*       1017     2036: contig of 1020 bp in length
*                      gap of unknown length
*       2037     3371: contig of 1335 bp in length
*                      gap of unknown length
*       3372     4972: contig of 1601 bp in length
*                      gap of unknown length
*       4973     5981: contig of 1009 bp in length
*                      gap of unknown length
*       5982     8008. contig of 2027 bp in length
*                      gap of unknown length
*       8009     9286: contig of 1278 bp in length
*                      gap of unknown length
*       9287    11243. contig of 1957 bp in length
*                      gap of unknown length
*      11244    14312: contig of 3069 bp in length
*                      gap of unknown length
*      14313    18423: contig of 4111 bp in length
*                      gap of unknown length
*      18424    23065. contig of 4642 bp in length
*                      gap of unknown length
*      23066    30558: contig of 7493 bp in length
*                      gap of unknown length
*      30559    37537: contig of 6979 bp in length
*                      gap of unknown length
*      37538    44831  contig of 7294 bp in length
*                      gap of unknown length
*      44832    50944: contig of 6113 bp in length
*                      gap of unknown length
*      50945    58263: contig of 7319 bp in length
*                      gap of unknown length
*      58264    65216: contig of 6953 bp in length
*                      gap of unknown length
*      65217    72310: contig of 7094 bp in length
*                      gap of unknown length
*      72311    78799. contig of 6489 bp in length
*                      gap of unknown length
*      78800    87607. contig of 8808 bp in length
*                      gap of unknown length
*      87608    98337. contig of 10730 bp in length
*                      gap of unknown length
*      98338   112198: contig of 13861 bp in length
*                      gap of unknown length
*     112199   126330: contig of 14132 bp in length
*                      gap of unknown length
*     126331   140617: contig of 14287 bp in length
*                      gap of unknown length
*     140618   153843. contig of 13226 bp in length
*                      gap of unknown length
*     153844   169798: contig of 15955 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a   34814 c   34487 g   49065 t      2 others
ORIGIN


alignment_scores:
            Quality:  407.00              Length:     77
              Ratio:  5.286                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-8 x AC009556   ..

 Align seq 1/1  to: AC009556  from: 1  to: 169798

      50 GluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLy 66
         ||||||||||||||||||||||||||  |  ||||||||||||||||||
  119825 GAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAGAAATCAA 119874

      66 sLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerI 83
         |||||||||||| ||||||||||||||||  ||||||||||||||||||
  119875 GCTCACAGCCCCAAAIATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 119924

      83 leLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVal 99
         |||||||||||||||||||||||||||||||||||||||||||||||||
  119925 TTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTG 119974

     100 AspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGlyHi 116
         |||| |||| |||| |||| |||| |||| |  |||||||||||||||
  119975 GATGAACTTGATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGGCCA 120024

     116 sGlyLysSerLeuSerTrpGlnGlyGlnSer 126
         ||||||||||||||||||||||||||||||
  120025 TGGTAAGTCACTAAGTTGGCAAGGCCAAAGT 120055

seq_name: gb_htg3:AC011488

seq_documentation_block:
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

     *          1     1007: contig of 1007 bp in length
     *                      gap of unknown length
     *       1008     1936: contig of 929 bp in length
     *                      gap of unknown length
     *       1937     2810: contig of 874 bp in length
     *                      gap of unknown length
     *       2811     3791: contig of 981 bp in length
     *                      gap of unknown length
     *       3792     4639: contig of 848 bp in length
     *                      gap of unknown length
     *       4640     5254: contig of 615 bp in length
     *                      gap of unknown length
     *       5255     5713: contig of 459 bp in length
     *                      gap of unknown length
     *       5714     6345: contig of 632 bp in length
     *                      gap of unknown length
     *       6346     6971: contig of 626 bp in length
     *                      gap of unknown length
     *       6972     7837: contig of 866 bp in length
     *                      gap of unknown length
     *       7838     8339: contig of 502 bp in length
     *                      gap of unknown length
     *       8340     8880: contig of 541 bp in length
     *                      gap of unknown length
     *       8881     9498: contig of 618 bp in length
     *                      gap of unknown length
     *       9499    10379: contig of 881 bp in length
     *                      gap of unknown length
     *      10380    10992: contig of 613 bp in length
     *                      gap of unknown length
     *      10993    11901: contig of 909 bp in length
     *                      gap of unknown length
     *      11902    13352: contig of 1451 bp in length
     *                      gap of unknown length
     *      13353    13947: contig of 595 bp in length
     *                      gap of unknown length
     *      13948    15016: contig of 1069 bp in length
     *                      gap of unknown length
     *      15017    15614: contig of 598 bp in length
     *                      gap of unknown length
     *      15615    16687: contig of 1073 bp in length
     *                      gap of unknown length
     *      16688    18152: contig of 1465 bp in length
     *                      gap of unknown length
     *      18153    19221: contig of 1069 bp in length
     *                      gap of unknown length
     *      19222    20477: contig of 1256 bp in length
     *                      gap of unknown length
     *      20478    21388: contig of 911 bp in length
     *                      gap of unknown length
     *      21389    22058: contig of 670 bp in length
     *                      gap of unknown length
     *      22059    23154: contig of 1096 bp in length
     *                      gap of unknown length
     *      23155    23925: contig of 771 bp in length
     *                      gap of unknown length
     *      23926    24290: contig of 365 bp in length
     *                      gap of unknown length
     *      24291    24907: contig of 617 bp in length
     *                      gap of unknown length
     *      24908    25839: contig of 932 bp in length
     *                      gap of unknown length
     *      25840    27116: contig of 1277 bp in length
     *                      gap of unknown length
     *      27117    27904: contig of 788 bp in length
     *                      gap of unknown length
     *      27905    28519: contig of 615 bp in length
     *                      gap of unknown length
     *      28520    28930: contig of 411 bp in length
     *                      gap of unknown length
     *      28931    30384: contig of 1454 bp in length
     *                      gap of unknown length
     *      30385    31245: contig of 861 bp in length
```

```
            *                    gap of unknown length
            *    31246    32149: contig of 903 bp in length
            *                    gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *                    gap of unknown length
            *    33959    34920: contig of 962 bp in length
            *                    gap of unknown length
            *    34921    34782: contig of 762 bp in length
            *                    gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *                    gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *                    gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *                    gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *                    gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *                    gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *                    gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *                    gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *                    gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *                    gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *                    gap of unknown length
            *    55099    81967: contig of 26790 bp in length
            *                    gap of unknown length
            *    81868   112000: contig of 30733 bp in length
            *                    gap of unknown length
            *   112001   179301: contig of 67301 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42598 t     85 others
ORIGIN


alignment_scores:
            Quality:  261.00             Length:     75
              Ratio:  3.783                Gaps:      0
 Percent Similarity:  92.000   Percent Identity: 60.000

alignment_block:
 US-09-174-937-8 x AC011488/rev   ..

 Align seq 1/1  to reverse of: AC011488  from: 1  to: 179301

      51 ProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLe 67
         ||||||||||::|||::::::|||   :::||    :::|||       ||
  124202 CCTCCCAACAGCTTAAATTTGAATGGAGCCCATTGTAAGAGGAAGACGCT 124252

      67 uThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIleL 84
         |:::|:|:::::|:|:|::::|:|||||:|| :|||||||||:|||:::|
  124251 GGTGGCCCAGAGATTAATATTTCTCTGGATCAGTCAGGGCTCCCTGC 124202

      84 ouSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnValAsp 100
         ||||||||||   ||||||:::||||||::::::|||||||||||||||
  124201 TCTCCGATGACTTCTTGGATACCCCTGATGACCTGGATATTAACGTGGAT 124152

     101 GluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGlyHisGl 117
         ::::::::|||||||||||||:::|||||:::::::|||   |||:::||
  124151 GACATCGACACCCCGGATGAGGCCGACTCCTTTGAGTTCACGGGGCACGG 124102

     117 yLysSerLeuSerTrpGlnGlyGln 125
         |:::::::::::|:::::|:::
  124101 CAACGAACTGGAGTGGGAAGGTAAA 124077
```

```
seq_name: gb_pat:AR028453

seq_documentation_block:
LOCUS       AR028453     1386 bp    DNA             PAT       24-SEP-1999
DEFINITION  Sequence 22 from patent US 5858678.
ACCESSION   AR028453
VERSION     AR028453.1  GI:5940447
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G
  TITLE     Apoptosis-regulating proteins
  JOURNAL   Patent: US 5858678-A 22 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    336 g    372 t
ORIGIN


alignment_scores:
            Quality:  147.00             Length:     89
              Ratio:  2.333                Gaps:      5
 Percent Similarity:  70.787   Percent Identity: 43.820

alignment_block:
 US-09-174-937-8 x AR028453

 Align seq 1/1  to  AR028453  from  1  to  1386

      27 GluAspValGlyMetAspIleProPheGlnGluGlyValLeuSerPro 46
         ||||||     ...   ...|||:  ::.....:::::   :::
     250 GAAGATTTTCCGATACCTTTACCAGAAGATGATAGTATTGAAGCAGATAT 299

      43 rAlaAlaAspMetArgProGlu  ProProAspSerLeuAspLeuAsnA 59
         |||         ||||||     |||:::|||||:::::|||
     300 ACTAGCTATAACTGGACCAGAGGACCAGCCTGGCTCACTAGAAGTTAATG 349

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         ::.    |||...  ||||||   ||||||.::|||:::|| :::
     350 GAAATAAAGTGAGAAAG...AAACTAATGGCTCCAGACATTAGCCTGATA 396

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   |||::::|||||:::::|||| |||||   |:'||:::
     397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT   TTGGATAAAAG 446

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         |||||||||:::      ||'  ||||||||||||:::||:  :
     444 TGGGGAGATTGACTTA......GATGGCTTAGACACACCGTCAGAGAATA 489

     108 laAspSerPheGluTyr 113
         ::::::::|||||||::
     488 GTAATGAGTTTGAGTGG 504

seq_name: gb_pr3:HSU15173

seq_documentation_block:
LOCUS       HSU15173     2382 bp    mRNA            PRI       [illegible]
DEFINITION  Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2)
            mRNA, complete cds
ACCESSION   U15173
VERSION     U15173.1  GI:558843
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 212 to 1156)
  AUTHORS   Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K.,
            Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.
```

```
  TITLE     Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
            of cellular proteins
  JOURNAL   Cell 79, 341-351 (1994)
  MEDLINE   95042730
REFERENCE   2  (bases 1 to 2382)
  AUTHORS   Boyd,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1994) Janice M. Boyd, Institute for Molecular
            Virology, St. Louis University Health Sciences Center, 3681 Park
            Avenue, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B-cell, S.Elledge; B-cell, J.Ambrus; fetal
                     brain, M.Green"
                     /cell_type="B cells; brain"
                     /dev_stage="fetus (brain)"
     gene            1. .2382
                     /gene="BNIP2"
     CDS             212. .1156
                     /gene="BNIP2"
                     /codon_start=1
                     /product="BCL2/adenovirus E1B 19kD-interacting protein 2"
                     /protein_id="AAC00021.1"
                     /db_xref="GI:558844"
                     /translation="MEGVELKEEWQDEDFPIPLPEDDSIEADILAITGPEDQPGSLEV
                     NGNKVRKKLMAPDISLILDPSDGSVLSDDLDESGEIDLDGLDTPSENSNEFEWEDDLP
                     KPKTTEVIPKGSITEYTAAEEKEDGPPWFMFFYGEQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQPNYRYLMDNLFKYVIGTLELLVAENYMIVYLNGATIRR
                     KMFSLGWLPKCYQQIDFELPENLKSLIIVHPSWFIRTLLAVTRPFISSKFSQKIRYVF
                     NLAELAELVPMEYVGIPECIKQVDQELNGKQDEPKNEQ"
BASE COUNT      734 a    393 c    511 g    744 t
ORIGIN


alignment_scores:
            Quality:  147.00             Length:     89
              Ratio:  2.393                Gaps:      5
 Percent Similarity:  70.787   Percent Identity: 43.820

alignment_block:
 US-09-174-937-8 x HSU15173   ..

 Align seq 1/1  to: HSU15173  from: 1  to: 2382

      27 GluAspValGlyMetAspIleProPheGluGlyGlyValLeuSerProSe 43
         ||||||     :::   :::|||  ::::::::::::   :::
     248 GAAGATTTTCCGATACCTTTACCAGAAGATGATAGTATTGAAGCAGATAT 297

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         |||       |||||||      || :::|||||||:::::|||
     298 ACTAGCTATAACTGGACCAGAGGACCAGCCTGCCTCACTAGAAGTTAATG 347

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||:::   ||||||   ||:||:···|||···||:::
     348 GAAATAAAGTGAGAAAG...AAACTAATGGCTCCAGACATTAGCCTGACA 394

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   |||:::|||||||:::|||||||| ||   ||||||:::
     395 CTGGATCCTAGTGAAGGTTCTGTAATTGTCAGATGAT...TTGGATGAAAG 441

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         |||||||||:::     |||  || |:|||||||:::|||   :
     442 TGGGGAGATTGACTTA.    .GATGGCTTAGACACACCGTCAGAGAATA 485

     108 laAspSerPheGluTyr 113
         ::::::::|||||:::
     486 GTAATGAGTTTGAGTGG 502

seq_name: gb_rc:AF035207

seq_documentation_block:
```

```
LOCUS       AF035207     2213 bp    mRNA            BCT       25-FEB-1998
DEFINITION  Mus musculus Nip21 mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip21 gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, Incyte Pharmaceuticals,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1. .2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="between D2Mit59 and D2Mit60"
     gene            1. .2213
                     /gene="Nip21"
     CDS             134. .1114
                     /gene="Nip21"
                     /note="associates with motifs of Bcl2 type adenovirus E1B
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP21"
                     /protein_id="AAC24829.1"
                     /db_xref="GI:2911348"
                     /translation="MEGVELKDEWQDEDFPIPLPEDDSIEADILAITGPEDQPGSLEV
                     NGNKVRKKLMAPDISLILDPSDGSVLSDDLDESGEIDLDGLDTPSENSNEFEWEDDLP
                     KPKTTEVIPKGSITEYTAAEEKEDGPPWFMFFYGEQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQPNYRYLMDNLFKYVIGTLELLVAENYMIVYLNGATIRR
                     KMFSLGWLPKCYQQIDFELPENLKSLIIVHPSWFIRTLLAVTRPFISSKFSQKIRYVF
                     NLAELAELVPMEYVGIPECIKQVDQELNGKQDEPKNEQ"
BASE COUNT      637 a    400 c    501 g    605 t
ORIGIN


alignment_scores:
            Quality:  135.00             Length:     89
              Ratio:  2.177                Gaps:      5
 Percent Similarity:  69.663   Percent Identity: 44.944

alignment_block:
 US-09-174-937-8 x AF035207   ..

 Align seq 1/1  to: AF035207  from: 1  to: 2213

      27 GluAspValGlyMetAspIleProPheGluGlyGlyValLeuSerProSe 43
         ||||||     :::   :::|||  ::::::::::::   :::
     170 GAAGATTTTCCAATACCTTTACCAGAAGATGATAGTATTGAAGCAGATAT 219

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :         ||:::      ::  ::: :  :||||:
     220 ACTAGATGGAACTGATCCAGACAGACATCCTGTTTCTTATAGAAGTTAATG 269

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||:::   ||||||   ||:||:···|||···||:::
     270 GGAACAAAGTAAGGAAG...AAATTGATGGCTCCAGACATTAGCCTGACA 316

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   |||:::|||||||:::|||||||| ||   ||||||:::
     317 CTGGATCCTGGTGAAGGTTCTGTAATTGTCAGATGAT...TTGGATGAAAG 366

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         |||:::|  |:::      :::                 : :      :
```

```
     364 TGGAGAGGTTGACCTG......GAAGGCTTAGATACCCCATCAGAGAACA 407

     108 IaAspSerPheGluTyr 113
         :::::::::::::::::
     408 GTGATCAGTTTGATGGG 424

seq_name: gb_htg6:AC009556

seq_documentation_block.
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6473159
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depante,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Hea ford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liew,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Mulla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pillara,V., Pilo,P., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,T., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yo,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: D1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13: M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *    1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *    2037     3371: contig of 1335 bp in length
            *                   gap of unknown length
            *    3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *    4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *    5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *    8009     9286: contig of 1278 bp in length
            *                   gap of unknown length
            *    9287    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *   11244    14312: contig of 3069 bp in length
            *                   gap of unknown length
            *   14313    18423: contig of 4111 bp in length
            *                   gap of unknown length
            *   18424    23065: contig of 4642 bp in length
            *                   gap of unknown length
            *   23066    30558: contig of 7493 bp in length
            *                   gap of unknown length
            *   30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *   37538    44831: contig of 7294 bp in length
            *                   gap of unknown length
            *   44832    50944: contig of 6113 bp in length
            *                   gap of unknown length
            *   50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *   58264    65216: contig of 6953 bp in length
            *                   gap of unknown length
            *   65217    72310: contig of 7094 bp in length
            *                   gap of unknown length
            *   72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *   78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *   87608    98337: contig of 10730 bp in length
            *                   gap of unknown length
            *   98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *  112199   126330: contig of 14132 bp in length
            *                   gap of unknown length
            *  126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *  140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *  153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a   34914 c   34187 g   49365 t      2 others
ORIGIN


alignment_scores:
            Quality:  118.00              Length:      39
              Ratio:  3.806                 Gaps:       1
 Percent Similarity:  79.487   Percent Identity: 69.231

alignment_block:
 US-09-174-937-8 x AC009556/rev   ..

 Align seq 1/1  to reverse of: AC009556  from: 1  to: 169798
```

```
      1 MetLeuLysSerCysSerArgAlaSerPheSe:ProSerValArgLysPr 17
    97933 ATGCTGAAGAGCTTGTAGTCGTGCATCCTTTCAACCCTCGGTTAGAAAGCC 97884

     17 oProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMetAspIleP 34
    97883 TCCTCTCATCCTCAGAAGG.............TAGGTGGGTCCTGGTGTAG 97846

     34 roPheGluGluGlyVal 39
    97845 CTTTTAGGAGATCGGTA 97829

seq_name: gb_htg5:AC009481

seq_documentation_block:
LOCUS       AC009481    192354 bp    DNA             HTG       17-NOV-1999
DEFINITION  Homo sapiens clone NH0303D03, *** SEQUENCING IN PROGRESS ***, 2
            unordered pieces.
ACCESSION   AC009481
VERSION     AC009481.2  GI:6447143
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192354)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192354)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 17, 1999 this sequence version replaced gi:5762586.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    84479  contig of 84479 bp in length
            *   84480    84497  gap of unknown length
            *   84498   192354  contig of 107857 bp in length.
FEATURES             Location/Qualifiers
     source          1..192354
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0303D03"
BASE COUNT    58604 a   39409 c   39230 g   55093 t     18 others
ORIGIN


alignment_scores:
            Quality:   91.50              Length:    115
              Ratio:    1.360               Gaps:      7
 Percent Similarity:   89.130   Percent Identity:  89.565

alignment_block:
 US-09-174-937-8 x AC009481/rev

 Align seg 1/1  to reverse of: AC009481  from: 1  to: 192354

       2 LeuLysSerCysSerArgAlaSerPheSerProSerVal..... 14
    92842 CTTGCGTCACTATCTTTTTCTCGCTTAGGCCTTTAGTCAAATTCTTTCT 92793

      15 ....ArgLysProProLeuIleLeuArgArgL:LeuSerGluAspValG 30
```

```
   92792 TCCAAGGAGACAAGAATTGAGATTTTGTA.....AATA 92758

      30 lyMetAspIleProPheGluGlyValProPro 41
   92757 GTATGGATTTGCCACTA.....TTAATTTTAATTTA 92721

      47 MetArgProGluProProAsnGlyAspProAsnAspThrPheAr 61
   92722 .....AAATATGGTAATTCTTTATTTCTTAAAATTTAAATTTTAT 92676

      63 gArgIleLysLeuThrAlaProAsnIleAsnLysSerLeuAspVa 78
   92681 AAAATGTAGCATTCCTATGTGTAATTTAATATGCTCTTAATA.... 92638

      80 luGlySerIleLeuSerAspAspAsnLeuAspProAsp..... 91
   92637 .....TTTTTAACAGAAGAATTTTTTCTTAATTTATT..... 92598

      95 IleAspIleAsnValAspGluLysAspThrArgAspIleAsnAsp 110
   92593 ATAGAC...AATTCTGATGATTCAATGGTTCTTATGATTTATTT 92557

seq_name: gb_pr2:HSU95738

seq_documentation_block:
LOCUS       HSU95738   171368 bp    DNA             PRI       [illegible]
DEFINITION  Human chromosome 16 [illegible] BAC clone [illegible]
            sequence.
ACCESSION   U95738
VERSION     U95738.1  GI:2338778
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; [illegible] Hominidae; Homo.
REFERENCE   1  (bases 1 to 171368)
  AUTHORS   Adams,M.D., Loftus,B., [illegible], Fuhrmann,C., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human chromosome 16p [illegible] BAC clone CIT987SK [illegible] complete sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171368)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-1996) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 171368)
  AUTHORS   Adams,M.D., Loftus,B., [illegible] Fuhrmann,C., Kim,U.J. and
            Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1996) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Aug 21, 1997 this sequence version replaced U95738.1.
            BAC clone CIT987SK-2-B12 of [illegible] chromosome
            16. Genes were identified by computational [illegible] X
            GRAIL (available by anonymous ftp from [illegible]),
            Genefinder (available by anonymous ftp [illegible]
            colin@u.washington.edu), GENSCAN (available via [illegible] e-mail server
            at genscan@genome.stanford.edu, [illegible] database at
            [illegible] http://www.[illegible] against a
            peptide database. Regions with [illegible] Kyte, J.,
            Kitchwell, B., Degma, [illegible] the
            identification and elimination of repetitive elements in DNA
            sequences. Computers Chem. 20:1-19(1996)) available by
            anonymous ftp from [illegible]
FEATURES             Location/Qualifiers
     source          1..171368
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.11"
                     /clone="29B12"
     repeat_region   2512..3611
                     /rpt_family="L1P"
```

```
repeat_region   2612. .2899
                /rpt_family="LTR7"
repeat_region   3354. .3690
                /rpt_family="SVA"
repeat_region   complement(3383. .3677)
                /rpt_family="ALU"
repeat_region   3697. .3958
                /rpt_family="MER4B"
repeat_region   complement(4261. .4548)
                /rpt_family="ALU"
repeat_region   5420. .5931
                /rpt_family="MLT2A1"
repeat_region   complement(6602. .6715)
                /rpt_family="SVA"
repeat_region   6603. .6889
                /rpt_family="ALU"
repeat_region   complement(6773. .6884)
                /rpt_family="SVA"
repeat_region   complement(8091   8347)
                /rpt_family="SVA"
repeat_region   9097. .9387
                /rpt_family="ALU"
repeat_region   complement(11099. .11390)
                /rpt_family="ALU"
repeat_region   11111. .11493
                /rpt_family="SVA"
repeat_region   complement(12587. .12949)
                /rpt_family="SVA"
repeat_region   12662. .12961
                /rpt_family="ALU"
repeat_region   complement(15177. .15360)
                /rpt_family="MLT2A1"
repeat_region   complement(15412   15469)
                /rpt_family="MLT2A1"
repeat_region   complement(15940. .16236)
                /rpt_family="ALU"
repeat_region   16618   16788
                /rpt_family="L1MA9"
repeat_region   complement(18780. .19046)
                /rpt_family="MIR"
repeat_region   complement(28732. .28965)
                /rpt_family="MLT1E"
repeat_region   complement(31245. .31524)
                /rpt_family="MSTA"
repeat_region   complement(31550. .31652)
                /rpt_family="MSTA"
repeat_region   33701. .33832
                /rpt_family="MER4A"
repeat_region   33870. .33942
                /rpt_family="MER4A"
repeat_region   33982. .34239
                /rpt_family="MER4B"
repeat_region   complement(35264. .35400)
                /rpt_family="MIR2"
repeat_region   complement(39145. .39493)
                /rpt_family="SVA"
repeat_region   39218. .39508
                /rpt_family="ALU"
repeat_region   41006. .41083
                /rpt_family="MIR2"
repeat_region   complement(42785. .42896)
                /rpt_family="MER2B"
repeat_region   complement(43474. .43761)
                /rpt_family="MIR2"
repeat_region   complement(44137. .44235)
                /rpt_family="MIR"
repeat_region   complement(45667. .45936)
                /rpt_family="ALU"
repeat_region   complement(46929. .47235)
                /rpt_family="ALU"
repeat_region   complement(53787. .54046)
                /rpt_family="SVA"
repeat_region   53791. .54077
                /rpt_family="ALU"
repeat_region   complement(55312. .55586)
                /rpt_family="MSTA"
repeat_region   complement(55588. .55731)
                /rpt_family="MSTA"
repeat_region   complement(60351. .60637)
                /rpt_family="ALU"
repeat_region   60365. .60648
                /rpt_family="SVA"
repeat_region   complement(61902. .62259)
                /rpt_family="MSTA"
repeat_region   70828. .71200
                /rpt_family="MLT1B"
repeat_region   74125   74411
                /rpt_family="ALU"
repeat_region   complement(76536. .76757)
                /rpt_family="MER30"
repeat_region   complement(78851. .79000)
                /rpt_family="MIR"
repeat_region   complement(81521. .81677)
                /rpt_family="MIR"
repeat_region   84733. .85063
                /rpt_family="MER44"
repeat_region   86087. .86249
                /rpt_family="L1MB3"
repeat_region   86250. .86934
                /rpt_family="L1ME3A"
repeat_region   87233. .87392
                /rpt_family="MER42C"
repeat_region   87481   87772
                /rpt_family="MER42A"
repeat_region   87774. .88306
                /rpt_family="MER42A"
repeat_region   88654   88995
                /rpt_family="MER42A"
repeat_region   complement(90553. .90976)
                /rpt_family="MER41"
repeat_region   91135. .91278
                /rpt_family="MSTA"
repeat_region   91280. .91550
                /rpt_family="MSTA"
repeat_region   complement(94311. .94541)
                /rpt_family="MSTA"
repeat_region   complement(94542   94678)
                /rpt_family="THE1B"
repeat_region   96186. .96346
                /rpt_family="MLT1A"
repeat_region   96372   96568
                /rpt_family="MLT1A"
repeat_region   102884. .103047
                /rpt_family="L1"
repeat_region   103048. .103451
                /rpt_family="L1"
repeat_region   103452   103805
                /rpt_family="L1PA2"
repeat_region   110262. .110543
                /rpt_family="MLT1F"
repeat_region   complement(113164. .113325)
                /rpt_family="MIR"
repeat_region   complement(113190   113290)
                /rpt_family="MIR"
repeat_region   complement(113562   113580)
                /rpt_family="MIR"
repeat_region   complement(123601. .123761)
                /rpt_family="MIR"
repeat_region   complement(124311. .124515)
                /rpt_family="MIR"
repeat_region   complement(127357. .127646)
                /rpt_family="ALU"
repeat_region   127371. .127630
                /rpt_family="SVA"
repeat_region   complement(138269. .138568)
                /rpt_family="ALU"
```

```
     repeat_region   138285. .138570
                     /rpt_family="SVA"
     repeat_region   complement(138925. .139204)
                     /rpt_family="SVA"
     repeat_region   138928. .139220
                     /rpt_family="ALU"

alignment_scores:
            Quality:   89.50            Length:    115
              Ratio:  1.467              Gaps:      5
 Percent Similarity:  53.043  Percent Identity: 27.826

alignment_block:
 US-09-174-937-8 x HSU95738/rev  ..

 Align seq 1/1  to reverse of: HSU95738  from: 1  to: 171468

      17 ProProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMet..... 31
         |||  |||:::|||:::::  |||:: ::::  |||:::
  114761 CCTTCACTGTTCCTTGAGAAAGAACTTGAAICCTCCCAGGGAITAGGAGG 114712

      32 ... AspIleProPheGluGluGlyValLeuSerProSerAlaAlaAspM 47
             ...|||  :::::::::  ||  :::::::|||
  114711 TCACCCACTGCCAGGGCAGTCTCCCCCACIGCTGGTAGATCTGTTGATG 114662

      47 etArgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArg 63
         ...|||...  :::          ||||:::|||
  114661 CTGAACCTAAAICTGCTTCC  .. . .  ....ACATATCCCACC 114630

      64 ArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGl 80
         :::   |||  |||
  114629 CCICTTCTGCCCACACTICCT.  ...  .  ...  ......... 114609

      80 uGlySerIleLeuSerAspAspAsnLeuAspSerProAspGluIleAspI 97
             |||       |||  :::|||  .|||. . ....
  114608 .  .GCCTCCCCCCACCCACTCCATCCCCCTGAGGAGCTAAATG 114569

      97 leAsnValAspGluLeuAspThrProAspGluAlaAspSerPheGluTyr 113
         :::::::::|'|||||  ||||||  :::|||'|  :::|||
  114568 IGAAAACIGATGAAGCIGACACIGIIGAGGAACCC..AAGTTTCTTCTG 114522

     114 ThrGlyHisGlyLysSerLeuSer . .. TrpGlnGlyGlnSer 126
         |||:::  |||  ||||||||||     |||:::|||  |||
  114521 ACAAACACAGGAGCCAGCIIATCCIIGGGCICCAGGGGCCTGTCC 114477

Seq_name: gb_pr3:HS357I16

seq_documentation_block:
LOCUS       HS357I16   134506 bp    DNA             PRI       23-NOV-1999
DEFINITION  Homo sapiens DNA sequence from PAC 357I16 on chromosome 1p36.13.
            Contains GSSs, genomic marker D1S449 and a CA repeat polymorphism,
            complete sequence.
ACCESSION   AL021921
VERSION     AL021921.1  GI:3171879
KEYWORDS    HTG; CA repeat polymorphism; D1S449.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134506)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-1998) sanger.ac.uk/HGP/Chr1/) Sanger Centre,
            Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery?sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 2, 1998 this sequence version replaced gi:2909408.
            IMPORTANT: This sequence is the entire insert of clone 357I16.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variations annotated may not be found in the sequence submission
            corresponding to the overlapping clone as we submit sequences with
            only a small overlap as described above.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            mapping group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1/.
            This sequence has been finished according to sequence set criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 357I16 is at 1 in this sequence. The
            true right end of clone 357I16 is at 134506.
            357I16 is from the library 4PR3 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong.
            For further details see http://bacpac.med.buffalo.edu/.
FEATURES             Location/Qualifiers
     source          1. .134506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.13"
                     /clone="RP3-357I16"
                     /clone_lib="RPCI-3"
     repeat_region   1. .227
                     /note="L1MC3 repeat: matches 226 .. 1245 of consensus"
     repeat_region   1597. .1988
                     /note="MIT1A2 repeat: matches 1 .. 374 of consensus"
     repeat_region   2061. .2461
                     /note="L1 repeat: matches 3331 .. 3708 of consensus"
     repeat_region   2514. .3887
                     /note="L1 repeat: matches 4112 .. 5534 of consensus"
     repeat_region   3738. .4632
                     /note="L1PA14 repeat: matches 1 .. 896 of consensus"
     repeat_region   4660. .5033
                     /note="L1 repeat: matches 3881 .. 4273 of consensus"
     repeat_region   5194. .5822
                     /note="SVA repeat: matches 4 .. 634 of consensus"
     repeat_region   6698. .6995
                     /note="AluSx repeat: matches 245 .. 542 of consensus"
     repeat_region   7355. .7958
                     /note="AluSg repeat: matches 1 .. 619 of consensus"
     repeat_region   7670. .7958
                     /note="L1 repeat: matches 4694 .. 5019 of consensus"
     repeat_region   8195. .8386
                     /note="L1 repeat: matches 5184 .. 5394 of consensus"
     repeat_region   8256. .8760
                     /note="L1MA8 repeat: matches 5 .. 573 of consensus"
     repeat_region   8406. .8985
                     /note="L1ME1 repeat: matches 111 .. 891 of consensus"
     misc_feature    10551. .10842
                     /note="match: GSS B71048"
     repeat_region   11647. .11840
                     /note="MER31 repeat: matches 1 .. 195 of consensus"
     repeat_region   11881. .11946
                     /note="MER31 repeat: matches 247 .. 311 of consensus"
     repeat_region   12089. .12165
                     /note="26 copies of 3 mer 85 % conserved"
     repeat_region   12167. .12378
                     /note="MIR repeat: matches 261 .. 441 of consensus"
     repeat_region   13033. .13174
                     /note="2 copies of 71 mer 91 % conserved"
     repeat_region   13104. .13341
                     /note="56 copies of 3 mer 85 % conserved"
     repeat_region   13154. .13673
                     /note="2 copies of 259 mer 87 % conserved"
     repeat_region   13256. .13548
                     /note="2 copies of 145 mer 82 % conserved"
     repeat_region   13415. .13642
                     /note="76 copies of 3 mer 84 % conserved"
     repeat_region   13876. .14021
                     /note="MIR2 repeat: matches 1 .. 146 of consensus"
     repeat_region   14234. .14349
                     /note="MIR repeat: matches 24 .. 154 of consensus"
```

```
repeat_region   15727. .15801
                /note="MIR repeat: matches 135. .59 of consensus"
repeat_region   15997. .16058
                /note="MIR repeat: matches 142. .79 of consensus"
repeat_region   16945. .17016
                /note="36 copies of 2 mer 86 % conserved"
repeat_region   16946   17011
                /note="3 copies of 22 mer 88 % conserved"
repeat_region   17500. .18053
                /note="L1ME1 repeat: matches 515. .2 of consensus"
repeat_region   18283. .18584
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   18955. .19110
                /note="MIR repeat: matches 206. .32 of consensus"
repeat_region   20059. .20256
                /note="MIR repeat: matches 27. .243 of consensus"
repeat_region   21415. .21618
                /note="MIR repeat: matches 28. .249 of consensus"
repeat_region   22164. .22419
                /note="MIR repeat: matches 78. .261 of consensus"
repeat_region   22258   22567
                /note="MIR repeat: matches 58. .262 of consensus"
repeat_region   26738. .26742
                /note="MIR repeat: matches 106. .144 of consensus"
repeat_region   26805. .26923
                /note="MIR repeat: matches 28. .153 of consensus"
repeat_region   27223. .27332
                /note="MER5A repeat: matches 47   163 of consensus"
repeat_region   27646. .27699
                /note="MIR2 repeat: matches 95. .146 of consensus"
repeat_region   28662. .28739
                /note="MIR repeat: matches 57   139 of consensus"
repeat_region   28903. .29031
                /note="MIR2 repeat: matches 2. .141 of consensus"
repeat_region   29524. .29581
                /note="29 copies of 2 mer 83 % conserved"
repeat_region   29685. .29748
                /note="MIR2 repeat: matches 82. .18 of consensus"
repeat_region   31781. .31866
                /note="MIR repeat: matches 58. .144 of consensus"
repeat_region   32630. .32766
                /note="MIR repeat: matches 52. .192 of consensus"
repeat_region   33390. .33483
                /note="MIR repeat: matches 20. .111 of consensus"
repeat_region   33838. .34074
                /note="MIR repeat: matches 16   153 of consensus"
repeat_region   34724. .34815
                /note="MIR repeat: matches 111. .22 of consensus"
repeat_region   35026. .35067
                /note="21 copies of 2 mer 93 % conserved"
repeat_region   35379. .35949
                /note="MLT1E repeat: matches 568. .1 of consensus"
repeat_region   35986. .36141
                /note="MIR repeat: matches 81. .252 of consensus"
repeat_region   36650   36744
                /note="MIR repeat: matches 142. .47 of consensus"
repeat_region   37515. .37565
                /note="3 copies of 17 mer 94 % conserved"
repeat_region   41761. .41805
                /note="23 copies of 2 mer 80 % conserved"
repeat_region   42019. .42319
                /note="L1PA2 repeat: matches 991   1 of consensus"
repeat_region   42770. .43214
                /note="L1 repeat: matches 5390. .5147 of consensus"
repeat_region   47697. .47873
                /note="MIR repeat: matches 260   84 of consensus"
repeat_region   48276. .48567
                /note="MIR repeat: matches 13   260 of consensus"
repeat_region   49303. .49735
                /note="MLT1A2 repeat: matches 374. .1 of consensus"
repeat_region   49837. .50005
                /note="MIR repeat: matches 262. .91 of consensus"
unsure          50754. .50925
                /note="forced join: verified by PAC PCR and pul
                information"
repeat_region   50907. .51072
                /note="L1ME3A repeat: matches 527. .362 of consensus"
repeat_region   51305. .51672
                /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region   53193. .53568
                /note="L1 repeat: matches 4226   3832 of consensus"
repeat_region   54179. .54212
                /note="MIR repeat: matches 102. .135 of consensus"
repeat_region   54433. .54848
                /note="MSTA repeat: matches 426. .1 of consensus"
misc_feature    56625. .56998
                /note="match: CSS B50226"
repeat_region   56988   57326
                /note="L1MC3 repeat: matches 2487. .2115 of consensus"
repeat_region   57676. .59900
                /note="L1MC3 repeat: matches 2189. .1 of consensus"
repeat_region   59763   60281
                /note="L1 repeat: matches 5390. .4841 of consensus"
repeat_region   60897   61246
                /note="MER41A repeat: matches 333. .3 of consensus"
repeat_region   61971   62219
                /note="MLT2_internal repeat: matches 2016. .1759 of
                consensus"
repeat_region   62239. .63072
                /note="MLT2_internal repeat: matches 1995. .1 of
                consensus"
repeat_region   63164. .63202

alignment_scores:
            Quality:   89.50              Length:    115
              Ratio:  1.467                Gaps:      5
 Percent Similarity:  53.043   Percent Identity: 27.826

alignment_block:
 US-09-174-937-8 x HS357I16   ..

 Align seq 1/1  to: HS357I16  from: 1  to: 134506

      17 ProProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMet..... 41
         |||   ||||:|||||:::::   ||| :: :::::   ||||::
   31169 CCTCCACTGTTCCTTGAGAAAGAACTTGAACCCTCCCAGGGATTAGGAGG 31218

      32 ...AspIleProPheGluGluGlyValIleSerProSerAlaAlaAspM 67
               :::|||     :  :.::   |||         ::::::||
   31219 ICACCCACTGCCAGGGCAGTCTGCCTTACTTGTTGGTAGATCCTGTTGATT 31268

      47 etArgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArg 86
         ...|||...      :::                  |||:::|'
   31269 CTGAACCTAAATCTGCTTCC.....................ACATATCCCAGG 31311

      64 ArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGl 80
         :::      |||   |||
   31301 CCGCTTCTGCCCACACTTCCT                  ..............  31321

      80 uGlySerIleLeuSerAspAspAsnLeuAspSerProAspGluIleAspI 47
                |||         ||    :::||||:: . :::::::
   31322 . .........GCCTCCCCCCACCCACTTCATGCCCTGACTAGCTAAATG 31361

      97 leAsnValAspGluLeuAspThrProAspGluAlaAspSerPheGlTyr 114
         ||||||   ||||||    ::: | ||||     :::::
   31362 TGAAAACTGATGAACCTGACACTCCTGATGAACCC   AATTTTCTTTTT 31411

     114 ThrGlyHisGlyLysSerLeuSer.....TrpGlnGlyGlnSer 128
         |||:::   |||   ||||||||      |||::|||    ::
   31409 ACAAACACAGGAGCCAGCTTATCTTTTTTCCTGGGGCTTCTCC 31457

seq_name. gb_ba1:AB017508

seq_documentation_block:
LOCUS       AB017508    [illegible] bp    DNA             BCT       [illegible]
DEFINITION  Bacillus halodurans C-[illegible] genomic DNA, [illegible] kb fragment, complete
```

```
            cds.
ACCESSION   AB017508
VERSION     AB017508.1  GI:4512395
KEYWORDS    .
SOURCE      Bacillus halodurans (strain C-125) DNA.
  ORGANISM  Bacillus halodurans
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (sites)
  AUTHORS   Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
            Horikoshi,K.
  TITLE     Sequence analysis of a 32-kb region including the major ribosomal
            protein gene clusters from alkaliphilic Bacillus sp. strain C-125
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
  MEDLINE   99206016
REFERENCE   2  (bases 1 to 32050)
  AUTHORS   Takami,H., Takaki,Y. and Nakasone,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideto
            Takami, Japan Marine Science and Technology Center, Deep-sea
            Microorganisms Research Group, 2-15 Natsushima, Yokosuka 237-0061,
            Japan (E-mail:takamih@jamstec.go.jp, Tel:0468-67-3895,
            Fax:0468-66-6364)
FEATURES             Location/Qualifiers
     source          1. .32050
                     /organism="Bacillus halodurans"
                     /strain="C-125"
                     /db_xref="taxon:86665"
     RBS             164. .171
                     /gene="rpoB"
                     /note="putative"
     gene            164. .3719
                     /gene="rpoB"
     CDS             177. .3719
                     /gene="rpoB"
                     /note="rpoB homologue (identity of 87% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75263.1"
                     /db_xref="GI:4512396"
                     /translation="MTGQLIQVGRHFQRFSYARINEVLELPNLIEIQTASYQWFLDEG
                     LREMFQDISPIQDFIGNLVLEFIDYSLGEPKYPVDESKERDVTYAAPLRVKVRLINKE
                     IGEVKEQEVFMGDFPLMIDTGTFIINGAERVIVSQLVRSPSVYYSEKIDKNGKKGYTA
                     IVIPNRGAWLELETDAKDIVYVRIDRTRKIPVTVLLRALGFGSDQEIIDLLGEDEYLR
                     NILEKDNIDSAEKALLEIYERLRPGEPPTVENAKSLLDSPFFDPKRYDLANVGPYKVN
                     KKLHIKNRLFNQRLAETLIDPETGEVIAEEGSIIDRRTLDPILPYLENNVGFPIVRMS
                     GGVVEEDEVRLQSVKIYAPDDQDGEHVIRVIGNGLVEKEIKHITPADIIASINYFFNL
                     LHSVGGTDDIDHLGNRRLRSVGELLQNQFRIGLSRMERVVFERMSIQDPNSIIPQALI
                     NIRPVIASIKEFFGSSQLSQFMDQTNPLAELTHKPRLSALGPGGLTPEPAGFEVPDVH
                     YSHYGFMCPIEIPEGPNIGLINSLSIYAKVNEFGFMETPYRRVDPEIGKVTAQIDYLI
                     ADEEDNYVVAQANMKLAEDGSFIDENIIARFRGENIVVSRDPVDYMDVSPKQVVSAAT
                     SCIPFLENDDSNRALMGANMQRQAVPLLVPEAPIVGTGMEHVSAKDSGAAIVSKHRGI
                     VERVIAKEIWVRRLEEVDGKEIKGDLDKYRLQKFIRSNQGTSYNQRPIVKEGDVVEKR
                     EILADGPSMEKGEMALGRNVLVGFMTWEGYNYEDAIILSERLVKDDVYTSIHIEEYES
                     EARDTKLGPEEITRDIPNVGEDALRNLDEPGIIRVGAEVKDGDILVGKVTPKGVTELT
                     AEERLLHAIFGEKAREVRDISLRAPHGGDGIVLDVKIFNREDGDELPPGVNQLVRVYI
                     VQKPKIHEGDKMAGRHGNKGVISRILPEEDMPYLPDGTPIDIMLNPLGVPSRMNIGQV
                     LELHLGMAARRLGLHVASPVFDGASEEDVWATLEEAGMARDGKTILYDGRTGEPFDNR
                     VSVGIMYMIKLAHMVDDKLHARSTGPYSLVTQQPLGGKAQFGGQRFGEMEVWALEAYG
                     AAYILQEILIVKSDDVVGRVKIYEAIVKGENVPEPGVPESFKVLIKELQSLGMDVKML
                     SSTEEEIEMKELDDDEDQASDKLNLNIDSTESNV"
     gene            3793. .7429
                     /gene="rpoC"
     RBS             3793. .3800
                     /gene="rpoC"
                     /note="putative"
     CDS             3809. .7429
                     /gene="rpoC"
                     /note="rpoC homologue (identity of 85% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75264.1"
                     /db_xref="GI:4512397"
                     /translation="MVDFKIPEDVNLRTYLITDPDTGKEIDRLYDFGTCVGVCCEDE
                     IQEMVPLAIKTKKVLRDAIREEMGGRNISFAGVEALIEEASKSIASGEYLDKIGTLTA
                     GGEWMSTFIGIDELDVKNKRFGDFEIEPTELFAIGADANLEAIMLGLTNVDKSLGSNM
                     IAVSIPEAVMEEAVTIEPLKRKVKPDWLVLKGNVTGKLVPAGTGYNFYNQLVQRTIL
                     AAIKGKPDELVTLKENVTGKLVPAGTGMNEYINQLIVEEMDLLAVSIPEAVMEEAVT
                     IE"
     gene            7523. .7784
                     /gene="ybxF"
     RBS             7523. .7528
                     /gene="ybxF"
                     /note="putative"
     CDS             7536. .7784
                     /gene="ybxF"
                     /note="ybxF homologue (identity of 79% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75265.1"
                     /db_xref="GI:4512398"
                     /translation="MSVFKVQLATLDSTGPLQTAKLDLMVVLVSKKADFPSYVNK
                     VEAMASVKQIFIYVDSKKKLDKAQGIEVSAAIVALKK"
     gene            7884. .8320
                     /gene="rpsL"
     RBS             7884. .7891
                     /gene="rpsL"
                     /note="putative"
     CDS             7901. .8320
                     /gene="rpsL"
                     /note="rpsL homologue (identity of 94% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75266.1"
                     /db_xref="GI:4512399"
                     /translation="MATINQLVRKPRARKVAKSNVPALEACPQKRGVCTRVYTTTPK
                     KPNSALRKVARVRLTNGFEVTAYIPGIGHNLQEHSVVLIRGGRVKD
                     LPGVRYHIVRGALDTAGVKDRKQSRSKYGTKRPKEKK"
     gene            8351. .8837
                     /gene="rpsG"
     RBS             8351. .8359
                     /gene="rpsG"
                     /note="putative"
     CDS             8367. .8837
                     /gene="rpsG"
                     /note="rpsG homologue (identity of 94% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75267.1"
                     /db_xref="GI:4512400"
                     /translation="MVDFKIPVVREFVLIPFSTNVRDVGAMGIAQKILY
                     NAFELVRERECREELAVLGLKNMKYLAKRIAVVGNNVRVSEPRRILGLR
                     WLVNYSRIRGFKTWFEFANETLDAANMTGAAVKKSFTPKHMNDLANKAPAHYRW"
     gene            8929. .11025
                     /gene="fus"
     RBS             8929. .8935
                     /gene="fus"
                     /note="putative"
     CDS             8947. .11025
                     /gene="fus"
```

```
                     /note="fus homologue (identity of 87% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75268.1"
                     /db_xref="GI:4512401"
                     /translation="MARFFSLPNTRNIGIMAHIDAGKTTTTERILFYTGRIHKIGEIH
                     EGASQMDWMEQEQERGITITSAATTAQWKNNRINIIDTPGHVDFTVEVERSLRVLDGA
                     VAVLDAQSGVEPRTETVWRQATTYGVPRVVFVNKMDKTGADFLYSVSTLHDRLQANAH
                     PIQLPIGAEDNFEGIIDLVDMVAYFYEDDLGTRIEAKEIPDEYKEQAQEYHEKLVEAA
                     AELDEELMMKYLEGEELTKDELKAAIRKGTCNVEFYPVLCGSAFKNKGVQLMLDAVLD
                     YLPSPLDVPAIKGHVPDTEEEAVRKPGDDQPFAALAFKVMTDPYVGKLTFFRVYSGTL
                     DSGSYVKNSTKDKPERVGRIVLMHANHREEISTVYSGDIAAAVGLKDTSTGDTLCDEK
                     NLVILESMEFPEPVIHLSVEPKSKADQDKMGLALAKLAEEDPTFKTHTDEETGQTIIA
                     GMGELHLDIIVDRLRREFKVEANVGAPQVSYRETIPQAAQVEGKFVRQSGGRGQYGHV
                     WIEFSPNEEGAGFEFVNGIVGSVVPREYIPSVQAGLEEALENGLLAGYPVIDIKAKLF
                     DGSYHDVDSSEMAFKIAASMALKNAKSKCNPVLLEPMMKVEVVVPEEYMGDVMGDITS
                     RRGRVEGMEARGNAQVVKAFVPLAEMFGYATSLRSRTQGRGTYTMEFDHYEEVPKSIS
                     EEIIKKNSGE"
     gene            11133..12340
                     /gene="tufA"
     RBS             11133..11140
                     /gene="tufA"
                     /note="putative"
     CDS             11150..12340
                     /gene="tufA"
                     /note="tufA homologue (identity of 91% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75269.1"
                     /db_xref="GI:4512402"
                     /translation="MAKEKFDRSKTHANIGTIGHVDHGKTTLTAAITTVLAKEGGKGV
                     AMAYDAIDGAPEEERERGITISTAHVEYETDNRHYAHVDCPGHADYVKNMITGAAQMDG
                     GILVVSAADGPMPQTREHILLSRQVGVPYLVVFLNKCDMVDDEELLELVEMEVRDLLS
                     EYDFPGDDVPVIRGSALKALEGDAEWEEKIIELMAAVDDYIPTPERDTEKPFMMPVED
                     VFSITGRGTVATGRVERGQLNVGDEVEIIGLEEEAKKTTVTGVEMFRKLLDYAEAGDN
                     IGALLRGVSREEVQRGQVLAKPGTITPHTNFKAEVYVLSKEEGGRHTPFFSNYRPQFY
                     FRTTDVTGIIQLPDGVEMVMPGDNVEMIVELIAPIAIEEGIKFSIREGGRTVGAGVVA
                     SIQK"
     gene            13248..13571
                     /gene="rpsJ"
     RBS             13249..13255
                     /gene="rpsJ"
                     /note="putative"
     CDS             13263..13571
                     /gene="rpsJ"
                     /note="rpsJ homologue (identity of 98% to B. subtilis )"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA75270.1"
                     /db_xref="GI:4512403"
                     /translation="MAKQKIRIRLKAYDHRVLDQSAEKIVETAKRSGANVSGPIPLPT
                     EKSVYTILRAVHKYKDSREQFEMRTHKRLIDIVNPTPQIVDALMRLDLPSGVDIEIKL

alignment_scores:
            Quality:   88.50            Length:     92
              Ratio:  1.725               Gaps:      2
 Percent Similarity:  55.435   Percent Identity: 26.087

alignment_block:
 US-09-174-937-8 x AB017503/rev

 Align seq 1,1  *  reverse of  AB017503  1..  1  to  32052

       6 SerArgAlaSerPheSerProSerValArgLysProProLeuIleLeuAr 22
         |:|:::     :::::::||||||:.||  ...   ||||||...   ||
   18419 TCAAATTTAATGTACGAACGACATCGTTACTACCTCAAGTTCTAGAACCG 18370

      22 gArgLeu.....LysSerGluAspValGlyM 31
         |   :::          :: :::       |||:
   18379 GACGATAACCGCTTTACAACGTCACCGTTCTCAACAGCGCCCCCTGGTG 18330

      31 etAspIleProPheGluGluGlyValLeuSerProSerAlaAlaAspMet 47
         ..           |||...   ...   |||' |      |||   ...

   18329 TTGCTTGCTTGACAGAGCAAACGATAACGTTACCAATGTTTTTAAG.TT..tcA

      48 ArgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgAr 64
         |||||||||||||||:::::||      ::::::::::::       ||
   18279 CGTCCAGAACCACCTAGCACTTTAATGCAAAGCAATTCACGAGCACCAGA 18464

      64 gIleLysLeuThrAlaProAsnIleAsn.....; 74
         :::::::   |||      ::::::::        |
   18429 GTTATCAGCTACTTTTAAACGGCTTTCTTGTTGCATCATTCGATCTTACT 18480

      74 euSerLeuAspGlnSerGluGlySer 82
         ||      ||||:|:::::|||::::
   18379 TCCCTTCGGATCAAGACGATGGCAAT 18354

seq_name: gb_ro:AF090436

seq_documentation_block:
LOCUS       AF090436     2256 bp    mRNA            ROD       ··MAR-1999
DEFINITION  Mus musculus dachshund variant 1 (Dach) mRNA, complete cds.
ACCESSION   AF090436
VERSION     AF090436.1  GI:4322566
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Caubit,X., Thangarajah,R., Theil,T., Wirth,J., Nothwang,H.J.,
            Ruther,U. and Krauss,S.
  TITLE     Mouse Dac, a novel nuclear factor with homology to Drosophila
            dachshund shows a dynamic expression in the central nervous system, the eye,
            the neocortex, and the limb bud
  JOURNAL   Dev. Dyn. 214 (1), 66-80 (1999)
  MEDLINE   99112783
REFERENCE   2  (bases 1 to 2256)
  AUTHORS   Caubit,X., Thangarajah,R. and Krauss,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1998) Department of Molecular Genetics, Institute
            of Medical Biology, University of Tromso, Tromso 9037, Norway
FEATURES             Location/Qualifiers
     source          1..2256
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="limb bud"
     gene            1..2256
                     /gene="Dach"
     CDS             1..2256
                     /gene="Dach"
                     /function="putative transcription regulator"
                     /note="nuclear protein; similar to Drosophila melanogaster
                     dachshund protein; encodes protein with a conserved (FSI)
                     rich domain in NH2 terminal domain; encodes 3 motifs similar
                     to core domain of vertebrate ski and ski related proteins
                     (sno)"
                     /codon_start=1
                     /product="dachshund variant 1"
                     /protein_id="AAD16097.1"
                     /db_xref="GI:4322567"
                     /translation="MAVPAALIPPTQLVPPQPPISTSASSSHTTSTSSATSSPAPSI
                     GPPASSGPTLPPPPPATSASSSAAATVISPSSSSSSSSSSSSSSSSSSSSSSSSCGPL
                     LAAGSSGGGVSAGGGGASSIPITASTGSSSSSSSSSSSSSSSSSSSSSSSSSSSCGPL
                     PGKPVYSTPSPVPPNPQNPQFMVILPSAPVAPIVPIPIPLPPLPPLFLPPHLVTG
                     LHTVYTKLKRLEITPVVCNVEQVRILRGLGATQPGVNPCKLLSPKTPEILYNDCTNAS
                     SPPGPPPKPTQCVTSPENSHIMPHSVPTLMSPTIIPNPTPIAAAAAAAAIINAAIAEA
                     MKVKKIKLEAMSNYHASNNQHGADSENGDMNSSVGSPGGSWTKETIHSPPSQGSQAPV
                     AHAPMPAAFSLPVSPINPLPPSHLPFVSIPIPPMMMPPPPPPPPPPPPPVTMAMSQM
                     NHLSTIANMAAAAQVQSPPSRVETSVIKERVPDSPSPAPSLPPGRRISHPSSHRSSS
                     VSSSPARTESSSDRIPVHQNGISMNQMLMGLSPNVLPGPKPNDLAGHIMTHESKRIHI
                     EKDETPLSTPTARNSIDALSLIGHGQPLPPGSPAPLPLPLSSAPLLNIQSLLKV
                     AIDNARAQEKQVLEKTELKMELPLPEPLPETTPRLTMMSLANRAIVLPKIKKEKKAK
                     PKLQEALEFETKPPEQAPQTIRLAASAPLSIPVINQSVIPPIPAPPAPSSAPADAERTIQD
                     GPLYLKITTVMY"
BASE COUNT      565 a    678 c    590 g    423 t
```

```
ORIGIN


alignment_scores:
            Quality:   84.50              Length:    127
              Ratio:  1.243                Gaps:      6
 Percent Similarity:  53.543   Percent Identity: 27.559

alignment_block:
 US-09-174-937-8 x AF090436   ..

 Align seg 1/1  to: AF090436  from: 1  to: 2256

       4 SerCysSerArgAlaSerPheSerProSerValArgLysProProLeuIl 20
         :::|||||||   :::...   ...|||:::::::::::|||||:::::
    1458 GCCTGCTCCATCTCTGGAGGAGGGCCGGAGGTGGGCAGCCACCCATCCT 1507

      20 eLeuArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluG 37
         :                                          |||   ::::
    1508 C.......   ...  .   .  ...          ..ACACCGCAGCAGC 1521

      37 luGlyVal.LeuSerPro...    .  ...SerAlaAlaAspMetArgPr 49
         ::   |||   |||||||           |||::::::|||      ||
    1522 AGTGTGTCCAGCTCCCCGGGCGCGGACTGAGAGTTCTTCCGACAGAATCCC 1571

      49 oGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleL 66
         :          |||::|||::::::|||:::
    1572 TGTCCATCAGAATGGCCTGTCCATGAACCAG.................A 1603

      66 ysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySer 82
           |||   :::   :::   |||::          :::|||||:::
    1604 TGCTTATGGGTTTATCCCCAAATGTGCTTCCTGGGCCAAAGGAGGGGGAT 1653

      83 IleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVa 99
         :::   :::   |||          :::    ...|||   |||.::
    1654 TTGGCTGGTCATGACATGGGGCATGAGTCAAAAGGGATCCACATTGAAAA 1703

      99 lAspGlu......LeuAspThrPro.........AspGluAlaAspSerP 111
         ||||||    ||||::|||||||          |||::::   |||::::
    1704 AGATGAGACCCCACTTTCCACACCAACCGCAAGAGACAGCATGGACAAAC 1753

     111 heGluTyrThrGlyHisGlyLysSerLeu 120
         :::::   |||||||||||||::   |||
    1754 TTTCTCTAACTGGGCATGGACAACCACTA 1782

seq_name: gb_ro:AF090437

seq_documentation_block:
LOCUS       AF090437     2100 bp    mRNA            ROD       03-MAR-1999
DEFINITION  Mus musculus dachshund variant 2 (Dach) mRNA, complete cds.
ACCESSION   AF090437
VERSION     AF090437.1  GI:4322569
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2100)
  AUTHORS   Caubit,X., Thangarajah,R., Theil,T., Wirth,J., Nothwang,H.G.,
            Ruther,U. and Krauss,S.
  TITLE     Mouse Dac, a novel nuclear factor with homology to Drosophila
            dachshund shows a dynamic expression in the neural crest, the eye,
            the neocortex, and the limb bud
  JOURNAL   Dev. Dyn. 214 (1), 66-80 (1999)
  MEDLINE   99112783
REFERENCE   2  (bases 1 to 2100)
  AUTHORS   Caubit,X., Thangarajah,R. and Krauss,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1998) Department of Molecular Genetics, Institute
            of Medical Biology, University of Tromso, Tromso 9037, Norway
FEATURES             Location/Qualifiers
     source          1. .2100
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="limb bud"
     gene            1. .2100
                     /gene="Dach"
     CDS             1. .2100
                     /gene="Dach"
                     /function="putative transcription regulator"
                     /note="nuclear protein; similar to Drosophila melanogaster
                     dachshund protein; encodes protein with threonine (PST)
                     rich domain in NH2 terminal domain; opposes domain similar
                     to core domain of vertebrate ski and sno related proteins
                     (sno)"
                     /codon_start=1
                     /product="dachshund variant 2"
                     /protein_id="AAD45098.1"
                     /db_xref="GI:4322569"
                     /translation="MAVPAAIPPTQIVPSLPRIGTPMSSSIITTGTGSATSSPAPSI
                     GPPASSGPTLPRPEPIAVSASSSAAATVTSPCPQQITSPQPTSPQNGDGGSNGNPS
                     LAAGSSGGGVSAGGGGASSTPIIASTDSSSSSSSSSSSSSSSSSSSSSSSSSCGPL
                     PGKPVVSTPSPVENTPLNNDPMVLLRPRVMTPVNGTELSTPLARLLPLKHLVGG
                     LHTVVTPKLKPLPTTVNGNVQLVELPGLDAVLPQVNPRRIQPRLPTTLYNLGINAS
                     SRPGRPPKRILSVTSPENSHIMPHSVPGLMSI IIIIIIIAAAAAAAAAAINAAIAEA
                     MKVKKTKLEAMSNVMPVNQLMADPTNCVNVLTLLLPLPMNNNNNLLLVSLPPASVT
                     MAMSQMNHLSIIANMAAAQVLSPFSPVPLSVLKPFVTLSPFAPSLEEIRRPGSHPS
                     SHRSSSVSSSPARTESGSLRLVHLNCLSMNLMLMSLSFNVLPGPKRGTLAGHDMGHE
                     SKRIHIEKDETPLSTPTAPDSIDKLSLTGHSLPTPLSPPSTFLPPLDLSSIETLLTNI
                     QGLLKVAIDNASLPVNLLEMTELKMDFLPEIDLDTLEK.IMFLANAAIVLKKLK
                     KEKKAKRKLQPALEPETKRRLAELILKQAAPADSLPVLNLPLTLRIPADRSGGRADA
                     ERTIQDGRLYIKTIVMY"
BASE COUNT      533 a    622 c    556 g    389 t
ORIGIN


alignment_scores:
            Quality:   84.50              Length:    127
              Ratio:  1.243                Gaps:      6
 Percent Similarity:  53.543   Percent Identity: 27.559

alignment_block:
 US-09-174-937-8 x AF090437   ..

 Align seg 1/1  to: AF090437  from: 1  to: 2100

       4 SerCysSerArgAlaSerPheSerProSerVa.ArgLysPr.Pr.LeIl 2
         :::'|||'   :::::   :::   :::::   :::   ::::
    1302 GCCTGCTCCATCTCTGGAGGAGGGCCGGAGGTGGGCAGCCACCCATCCT 1351

      20 eLeuArgArgLeuLeuSerGluAspValGlyMetAspIleP.PheGlu.G 37
         :                                          |||   ::::
    1352 C.......   ...  .   .  ...          ..ACACCGCAGCAGC 1365

      37 luGlyVal.LeuSerPro.....    .  ...SerAlaAlaAspMetArgPr 49
         ::   |||   |||||||           |||::::::|||      ||
    1366 AGTGTGTCCAGCTCCCCGGGCGCGGACTGAGAGTTCTTCCGACAGAATCCC 1415

      49 oGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleL 66
         :          |||::|||::::::|||:::
    1416 TGTCCATCAGAATGGCCTGTCCATGAACCAG.................A 1467

      66 ysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySer 82
           |||   :::   :::   |||::          :::|||||:::
    1468 TGCTTATGGGTTTATCCCCAAATGTGCTTCCTGGGCCAAAGGAGGGGGAT 1517

      83 IleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVa 99
         :::   :::   |||          :::    ...|||   |||.::
    1498 TTGGCTGGTCATGACATGGGGCATGAGTCAAAAGGGATCCACATTGAAAA 1547

      99 lAspGlu......LeuAspThrPro.........AspGluAlaAspSerP 111
         ||||||    ||||::|||||||          |||::::   |||::::
    1548 AGATGAGACCCCACTTTCCACACCAACCGCAAGAGACAGCATGGACAAAC 1597

     111 heGluTyrThrGlyHisGlyLysSerLeu 121
         :::::   |||||||||||||::   |||
```

```
    1598 TTTCTCTAACTGGGCATGGACAACCACTA 1626

seq_name: gb_ro:AF102547

seq_documentation_block:
LOCUS       AF102547     2256 bp    mRNA            ROD       01-NOV-1999
DEFINITION  Mus musculus dachshund (Dach) mRNA, complete cds.
ACCESSION   AF102547
VERSION     AF102547.1  GI:6165975
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Kozmik,Z.
  TITLE     Molecular cloning, expression and chromosomal localization of the
            Drosophila dachshund gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2256)
  AUTHORS   Kozmik,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1998) Laboratory of Transcriptional Regulation,
            Institute of Molecular Genetics, Flemingovo 2, Prague 16637, Czech
            Republic
FEATURES             Location/Qualifiers
     source          1..2256
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..2256
                     /gene="Dach"
     CDS             1..2256
                     /gene="Dach"
                     /note="transcription factor"
                     /codon_start=1
                     /product="dachshund"
                     /protein_id="AAF04742.1"
                     /db_xref="GI:6165976"
                     /translation="MAVPAAL.PPTQLVPPQPPISTSASSSGTITSTSSATISSPAPSI
                     GPPASSGPTLFPPEPIASSASSSAAATVTSPGGGGGGSGGGGGSGGNGGGGGSNCNPS
                     LAAGSSGGGVSAGGGGASSIPITASTGSSSSSSSSSSSSSSSSSSSSSSSSSSCGPL
                     PGKPVYSTPSPVENTPQNNPCKMVDLPGAKVASFTVEGCELICLPQAFDLFLKHLVGG
                     LHTVYTKLKRLEITPVVCNVFQVPILPGLGAIQPGVNPCKLISPKDFETLYNDCTNAS
                     SRPGRPPPKRTQSVTSPENSHIMPHSVPGLMSPGIIPPTGLTAAAAAAAAATNAAIAEA
                     MKVKKIKLEAMSNYHASNNQHGADSENGDMNSSVGSSGGSWDKEILHSPPSQGSQAPV
                     AHAPMPAAFSLPVSHPLNHLQHSHLPPNGLFLPFMMMPHPLIPVSLPPASVTMAMSQM
                     NHLSTIANMAAAAQVQSPPSPVETSVIKEPVPDSPSPAPSLEEGRPPGSHPSSHPSSS
                     VSSSPARTESSSDRIPVHQNGLSMNQMLMGLSPNVLPGPKEGDLAGHDMGHESKPIHI
                     EKDETPISTPTARDSIDKLSLFGHQGPLPPGFPSPFLFPDGLSSIETLLTNIQGLLKV
                     AIDNARAQEKQVQLEKTELPYPFLPEPELPFTLEKQLAMEQKNPAIVQKPLKKEKKAK
                     RKIQEAIFFEIKPPEQAPGTLPQAASADSLPVLNDSLIPEIEADPSGGPADAEPTIQD
                     GRLYLKTTVMY"
BASE COUNT      564 a    679 c    590 g    423 t
ORIGIN


alignment_scores:
            Quality:   84.50            Length:    127
              Ratio:  1.243              Gaps:      6
 Percent Similarity:  58.543   Percent Identity: 27.559

alignment_block:
 US-09-174-937-8 x AF102547   ..

 Align Set 1,1  to  AF102547  from  1  to  2256

       4 SerCysSerArgAlaSerPheSerProSerValArgLysProProLeuI 20
         :::|||'|'  :::::: :::|||||::::::'|||...|||||||...
    1458 GCCTGCTCCAGTGCTGGAGGAGGGCCGGAGTC.GGCAGCCAVCCATCCT 1507

      20 eLeuArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluG 37
         :                                    |||   ::::
    1508 :                                  ..ACACCCCAGTAGT 1521

      37 luGlyVal.LeuSerPro.....SerAlaAlaAspMetArgPr 49
         ::   |||'  ||||||      ||||::::::|'       :
    1522 AGTGTGTCCAGCTCCCCGGCGCGGAGTGAGAGTTCTTCCGACAGAAATCCC 1571

      49 oGluProProAsnSerLeuAspLeuAspAspThrHisProArgArgIleL 66
         |         ||||:: |||    :||  :::
    1572 TGTCCATCAGAATGGCCTGTCCATGAACCAG      .  .          A 1615

      66 ysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGlnGlySer 82
           |||   :::   :::   ||||::            :::.|||| :::
    1604 TGCTTATGGGTTTATCCCCAAATGTGCTTCCTGGGCCAAAGGAGGGGGAT 1653

      83 IleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVa 99
         :::   :::  |||      :::      ::|||   |||:::
    1654 ITGGCTGGTCATGACATGGGGCATGAGTCAAAACGGATCGACATTGAAAA 17

      99 lAspGlu . .  LeuAspThrPro.........AspGluAlaAspSerP 111
         ||||||      ||||::|||||         ||||::    |::::
    1704 AGATGAGACCCACTTTCCACATTAACCCCAGAGACACTATTCACAAAC 175

     111 heGluTyrThrGlyHisGlyLysSerLeu 120
         .....  ||||||||||||:..  |||
    1754 TTTCTCTAACTGGGCATGGACAACCACTA 1782

seq_name: gb_ro:MMU5669

seq_documentation_block:
LOCUS       MMU5669      2467 bp    mRNA            ROD       [illegible]-1998
DEFINITION  Mus musculus mRNA for dachshund protein.
ACCESSION   AJ005669
VERSION     AJ005669.1  GI:3367708
KEYWORDS    dach gene; dachshund protein; ski proto-oncogene.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2467)
  AUTHORS   Hammond,K.L., Hanson,I.M., Brown,A.G., Lettice,L.A. and Hill,R.E.
  TITLE     Mammalian and Drosophila dachshund genes are related to the Ski
            proto-oncogene and are expressed in eye and limb
  JOURNAL   Mech. Dev. 74, 121-131 (1998)
REFERENCE   2  (bases 1 to 2467)
  AUTHORS   Hammond,K.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1998) Hammond K.L., Developmental Genetics,
            Medical Research Council Human Genetics Unit, Western General
            Hospital, Crewe Road, Edinburgh, EH4 2XU, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..2467
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="E11.5 embryonic 5' stretch ("Lambda")"
                     /dev_stage="E11.5"
     gene            65..1774
                     /gene="dach"
     CDS             65..1774
                     /gene="dach"
                     /codon_start=1
                     /product="dachshund protein"
                     /protein_id="CAA05665.1"
                     /db_xref="GI:3367709"
                     /db_xref="SPTREMBL:O88716"
                     /translation="MVDLRGAKVASFTVEGCELICLPQAFDLFLKHLVTGLHTVYTKL
                     KRLEITPVVCNVFQVPILPGLGAIQPGVNPCKLISPKDFETLYNDCTNASSRPGRPPPK
                     RTQSVTSPENSHIMPHSVPGLMSPGIIPPTGLTAAAAAAAAATNAAIAEAMKVKKIKL
                     EAMSNYHASNNQHGADSENGDMNSSVGSSGGSWDKEILHSPPSQGSQAPVTHARMPAA
                     FSLPVSHPLNHLQHSHLPPNGLFLPFMMMPHPLIPVSLPPASVTMAMSQMNHLSTIAN
                     MAAAAQVQSPPSPVETSVIKEPVPDSPSPAPSLEEGRPPGSHPSSHPSSSVSSSPART
                     ESSSDRIPVHQNGLSMNQMLMGLSPNVLPGPKEGDLAGHDMGHESKPIHIEKDETPLS
                     TPTARDSIDKLSLFGHQGPLPPGFPSPFLFPDGLSSIETLLTNIQGLLKVAIDNARAQ
                     EKQVQLEKTELPYPFLPEPELPFTLEKQLAMEQKNPAIVQKPLKKEKKAKRKIQEALE
                     FEIKPPEQAPGTLPQAASPSLPVLNDSLIPEIEADPSGGPADAEPTIQDGRLYLKTT
```

```
                    VMY"
BASE COUNT      675 a    635 c    588 g    568 t      1 others
ORIGIN


alignment_scores.
            Quality:   84.50             Length:    127
              Ratio:  1.243               Gaps:      6
 Percent Similarity:  53.543   Percent Identity: 27.559

alignment_block:
 US-09-174-937-8 x MMU5669   ..

 Align seq 1/1  to: MMU5669  from: 1  to: 2467

       4 SerCysSerArgAlaSerPheSerProSerValArgLysProProLeuIl 20
         :::|||||||  ......  .:::|||::    .||..:|||||:::::
     976 GCCTGCTCCAIGTCTGGAGGAGGGCCGGAAYCCCGGCAGCCACCCATCCT 1025

      20 eLeuArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluG 37
         :                                              |||   ::::
    1026 C..              ..      . . . .          .. ACACCGCAGCAGC 1039

      37 luGlyVal.LeuSerPro............SerAlaAlaAspMetArgPr 49
         ::   |||    ||||||            ||.:::::|||      ||
    1040 AGTGTGTCCAGCTCCCCGGCGCGGACTGAGAGTTCTTCCGACAGAATCCC 1089

      49 oGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleL 66
         |         |||:::|||:::::::||:::
    1090 TGTCCATCAGAATGGCCTGTCCATGAACCAG      ...   ......A 1121

      66 ysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySer 82
           |||   :::   :::   |||:::           :::||||||:::
    1122 TGCTTATGGTTTATCCCCAAATGTGCTTCCTGGGCCAAAGGAGGGGGAT 1171

      83 IleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVa 99
         :::   :::   |||         :::     .::|||   |||:::
    1172 TTGGCTGGTCATGACATGGGGCATGAGTCAAAATGGATCCACATTGAAAA 1221

      99 lAspGlu......LeuAspThrPro.........AspGluAlaAspSerP 111
         '|||||      |||:::||||||         |||:::   |||::::
    1222 AGATGAGACCCCACTTTCCACACCAACCCCAACAGACAGCATCGACAAAC 1271

     111 heGluTyrThrGlyHisGlyLysSerLeu 129
         :::::  ||||||||||||:::   |||
    1272 TTTCTCTAACTGGGCATGGACAACCGCTA 1300
```

OM of: US-09-174-937-8 [... Issued_Patents_NA * ... pfs

Date: Feb 7, 2000 4:09 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1999 Compugen Ltd

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_qspool/US09174937/runat_1020000_095109_13295/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=1...00 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -IHP_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=10000000 -USEP=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-174-937-8
Query length: 126
Database: Issued_Patents_NA:*
Database sequences: 207703
Database length: 57918364
Search time (sec): 56.270000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-095-22	-	147.00	306.00	1.9e-09	1386	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-804-12	-	74.00	126.91	11.93	4849	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-805-12	-	74.00	126.91	11.93	4849	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-883-1	+	73.00	124.03	17.41	4853	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-807-1	-	73.00	124.03	17.41	4853	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-922-4	+	71.50	114.19	34.75	3511	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-829-922-6	+	71.50	116.95	12.19	2740	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-804-1	-	71.00	116.92	49.21	6716	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-1-5	+	69.50	131.71	7.17	1135	!	
/cgn2_6/prodata/1/ina/5D_COMB.seq:US-08-470-379-10	-	69.00	93.30	895.78	35190	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-511-807-6	-	67.50	104.87	5.60	666	!	S
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-804-7	-	67.50	104.71	57.46	3512	!	
/cgn2_6/pt/data/1/ina/PCTUS1_COMB.seq:PCT-US93-11319-3	-	67.50	131.97	5.60	499	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]-12	-	67.00	121.33	19.05	1485	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-883-2	+	67.00	123.33	19.05	1485	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-922-2	+	67.00	123.33	19.05	1485	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-807-5	-	67.00	116.85	40.69	2685	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-922-5	+	66.50	116.06	54.94	2855	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-811-5	-	66.50	116.06	54.94	2855	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]-1	-	65.50	113.27	110.68	5909	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]A-1	-	65.50	114.89	55.18	2351	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-881-922-19	+	65.50	114.51	58.26	2434	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-804-19	-	65.50	114.51	58.26	2434	!	
/cgn2_6/pt/data/1/ina/PCTUS4_COMB.seq:PCT-US94-12444-1	-	65.50	109.12	115.04	5994	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]-24	-	65.00	109.91	107.45	3576	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-[illegible]B-7	-	65.00	99.96	385.35	8367	!	
/cgn2_6/prodata/1/ina/5D_COMB.seq:US-08-470-113-7	+	65.00	99.88	385.35	8367	!	
/cgn2_6/pt/data/1/ina/PCTUS4_COMB.seq:PCT-US94-14456-29	-	65.00	114.61	115.04	5994	!	
/cgn2_6/prodata/1/ina/5B_COMB.seq:US-08-[illegible]-469-1	-	64.50	100.19	370.08	7326	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]-715-1	-	64.50	94.05	612.88	12847	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-[illegible]7A-3	-	64.00	94.03	815.68	11601	!	
/cgn2_6/pt/data/1/ina/FL_COMB.seq:US-08-511-[illegible]7A-24	-	64.00	94.03	815.68	11601	!	
/cgn2_6/prodata/1/ina/5A_COMB.seq:US-08-[illegible]-709A-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]-[illegible]A-1	-	61.50	116.55	47.09	1368	!	
/cgn2_6/pt/data/1/ina/5A_COMB.seq:US-08-[illegible]-[illegible]A-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/prodata/1/ina/5A_COMB.seq:US-08-470-[illegible]-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]-[illegible]A-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-511-[illegible]A-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-470-[illegible]-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]A-24	-	63.50	116.26	47.12	1368	!	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]-[illegible]B-15	-	63.50	111.89	259.16	5101	!	

/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]	-	[illegible]	[illegible]	[illegible]	498.16	5	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]	-	[illegible]	[illegible]	[illegible]	498.16	5	
/cgn2_6/pt/data/1/ina/5B_COMB.seq:US-08-[illegible]	-	[illegible]	[illegible]	[illegible]	498.16	5	
/cgn2_6/pt/data/1/ina/5D_COMB.seq:US-08-[illegible]	-	[illegible]	[illegible]	[illegible]	476.35	69	

Alignments

seq_documentation_block:
; Sequence 22, Application US/09[illegible]
; Patent No. 5854078
; GENERAL INFORMATION:
; APPLICANT: Chinnadurai, Govindaswamy
; TITLE OF INVENTION: APOPTOSIS-REGULATING PROTEINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/[illegible]
; FILING DATE: 22-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mack, Susan J.
; REGISTRATION NUMBER: 30,951
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 214..1155
US-08-408-095-22

alignment_scores:
Quality: 147.00 Length: [illegible]
Ratio: 2.383 Gaps: [illegible]
Percent Similarity: 70.787 Percent Identity: 47.4%

alignment_block:
US-09-174-937-8 x US-08-408-095-22 ...

Align seq 2/1 to US-08-408-095-22 ... from 1 to [illegible]

```
     397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 443

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         |||||||||::      |||   |.|||||||||:..|||   .
     444 TGGGGAGATTGATTA......GATGGCTTAGATACACGGTCAGAGAATA 487

     108 laAspSerPheGluTyr 113
         :::::::|||||::
     488 GTAATGAGTTTGAGTGG 504

Seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq; US-08-540-804-13

seq_documentation_block:
; Sequence 13, Application US/08540804
; Patent No. 5919666
;  GENERAL INFORMATION:
;    APPLICANT:  Young, Richard A.
;    APPLICANT:  Koleske, Anthony J.
;    APPLICANT:  Thompson, Craig M.
;    APPLICANT:  Chao, David M.
;    TITLE OF INVENTION:  No. 5919666el Factors Which Modify Gene
;    TITLE OF INVENTION:  Transcription and Methods of Use Therefor
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/540,804
;      FILING DATE:  11-OCT-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/521,872
;      FILING DATE:  21-AUG-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,265
;      FILING DATE:  25-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI94-03A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4849 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  148..4407
US-08-540-804-13


alignment_scores:
            Quality:   74.00              Length:    130
              Ratio:  1.014                Gaps:      6
 Percent Similarity:  56.154   Percent Identity: 26.923

alignment_block:
 US-09-174-937-8 x US-08-540-804-13    ..

Align seq 1/1  to: US-08-540-804-13  from: 1  to: 4849

      10 PheSerProSerValArgLysProPro[e:      ..........IleLe 21
         |||  ||||||   ..:||||||   :::
    1792 TTTTTTCCCTCATCTGAAAAACCCAATATCCGTACGATGAGCCGGAAGTAC 1841

      21 uArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluGlnG 68
         :::||||||   ::::::          :::|||||||||:::::|''
    1842 AAAAAGATTAAATGGAAAGAGGAAATATTTGGATATTCCATACATGAAA 1891

      28 lyValLeuSerProSerAlaAlaAs;Me'A:gP:pGlnP: P: AsnSer 54
         :::|||      |||          ;'   (:|     '.'
    1892 TGACCTTGCCAACGAGTCCATTATATATGGACCCAGGTGCGCCA..... 1935

      55 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaPro.. 7
         |||   .....      ...  ||'|||   ::::::   ||||||
    1936 CTCCCTGTGGAAACACCCCGCGATAGACGCAAAAGTGTGTTCGCTCCACC 1985

      71 .AsnIleAsnLeuSerLeuAspGlnSer.................GlnG 97
         ::: : :::           ..::::                 :::
    1986 GAATTTTAACCCCATAATAGAAAACAATGTTGATAACAAATACAAATCTG 2035

      81 lySerIleLeuSerAspAspAsnL:.AspSerProAspGluIle..... 98
         |:.    .|||.    .  ||       :::  :::
    2036 GAGGGAAATTTTCTTTCAGTCCGTTCCAAAAGCAGGAAGCATTAAAGTTT 2085

      96 AspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPheG: 112
         |||||.....  ...||:::::   ::::||   ::::::  '.
    2086 GATATTTCTATGGCGGATCTTTCTAGCTCTGAAGAGGAAGCAGCATCAACA 2135

     112 uTyrThrGlyHisGly.............LysSerLeuSer 121
         |   :::|||   :::           |||||:|||:::
    2136 AGAGAACGGTAGCAGCGATGAGGATCTAAAGTCATTGAAC 2175

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq; US-08-218-265-1

seq_documentation_block:
; Sequence 13, Application US/08218265
; Patent No. 5922585
;  GENERAL INFORMATION:
;    APPLICANT:  Young, Richard A.
;    APPLICANT:  Koleske, Anthony J.
;    APPLICANT:  Thompson, Craig M.
;    TITLE OF INVENTION:  No. 5922585el Factors Which Modify Gene
;    TITLE OF INVENTION:  Transcription and Methods of Use Therefor
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/218,265
;      FILING DATE:  25-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI94-03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4849 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  149..4407
US-08-218-265-13


alignment_scores:
            Quality:   74.00            Length:    130
              Ratio:  1.014              Gaps:      6
 Percent Similarity:  56.154   Percent Identity: 26.923

alignment_block:
 US-09-174-937-8 x US-08-218-265-13   ..

 Align seq 1/1  to: US-08-218-265-13  from: 1  to: 4849

      10 PheSerProSerValArgLysProProLeu          IleLe 21
         ||| |||'|| :::'|||:  :::
    1792 TTTTTTCCCTCATCTGAAAAACCGAATATCCGTACGATGAGCGGAACTAC 1841

      21 ArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluGluG 38
         :::'|||| :::::: ::::  |||||||||||||||||
    1842 AAAAAGATTAAATGGAAAGAGGAAATATTTGGATATTCCGATAGATGAAA 1891

      38 lyValLeuSerProSerAlaAlaAspMetArgProGluProProAsnSer 54
         :::|'|    |||      ||| |  |||
    1892 TGACCTTGCCAACGAGTCCATTATATATGGACCCAGGTGCGCCA..... 1935

      55 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaPro.. 70
         ||  :::::      :::  ||||||  ::::::  ||||||
    1936 CTCCCTGTGGAAACACCCCCGCGATAGACGCAAAAGTGTGTTCGCTCCACT 1985

      71 .AsnIleAsnLeuSerLeuAspGlnSer          GluG 81
         '|'::::|'|      ::::::::::::          :::|
    1986 GAATTTTAACCCCATAATAGAAAACAATGTTGATAACAAATACAAATCTG 2035

      81 lySerIleLeuSerAspAspAsnLeuAspSerProAspGluIle       95
         | :::  ::::|'   :::  |||:::::  :::  :::
    2036 GAGGGAAATTTTCCTTCAGTCCGTTGCAAAAGGAGGAAGCATTAAACTTT 2085

      96 AspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPheGl 112
         '|'|'::::::  :::|'::::::  :::'|  ::::::  ||
    2086 GATATTTTTATGGCGGATTTTCTAGCTCTGAAGAGGAAGAGGATGAAGA 2135

     112 uTyrThrGlyHisGly.....LysSerLeuSer 121
         :::'''  :::       |' ' ':::|:::
    2136 AGAGAACGGTACCAGCGATGAGGATCTAAAGTCATTGAAC 2175

seq_name: /tgrid_ftp1.data/1_1/na/FB_1_ME_s/US-08-832-883-1

seq_documentation_block:
; Sequence 1, Application US/08832883
; Patent No. 5807681
;  GENERAL INFORMATION:
;    APPLICANT:  Giordano, Antonio
;    APPLICANT:  Baldi, Alphonso
;    TITLE OF INVENTION:  METHODS FOR THE DIAGNOSIS AND PROGNOSIS
;    TITLE OF INVENTION:  OF CANCER
;    NUMBER OF SEQUENCES:  115
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FEITEL, LINDA, LAVIANO & MINNI, P.C.
;      STREET:  Suite 1800 Two Penn Center Plaza
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.  Version #1.
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/  /
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A
;      REGISTRATION NUMBER:  30,
;      REFERENCE/DOCKET NUMBER:  9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-
;      TELEFAX:  (215) 568-5549
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4853 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..3489
US-08-832-883-1


alignment_scores:
            Quality:   73.00            Length:    1
              Ratio:  1.553              Gaps:      
 Percent Similarity:  41.964   Percent Identity: 27.

alignment_block:
 US-09-174-937-8 x US-08-832-883-1   ..

 Align seq 1/1  to: US-08-832-883-1  from: 1  to: 1

      34 ProPheGluGluGlyValLeuSerProSerAlaAlaAspMetArg
         | |   :::  |          ::::::  :::
      49 CCGGGCGCCCAGGGGTGCGCTAT.....AGGAGGCAGCAGCAGCA

      50 ProProAsnSerLeuAspLeuAsnAspThrHisProArgArg
         | |
      99 GCCCCCG.....

      67 euThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGlu
                                  :::       ::::::
     106 ..CCTCCCCCTCCCCCCCCCCCACTCCTCCNTNGCCCCAA.....

      84 LeuSerAspAsp.....AspTyrAspSerPro
         ::: | | |                 :::  :::
     148 ...GAGGATGACGGCGAGGCGGAAGATACTCCCCCCCCTATCCCCTA

      32 G.....AspSerIle.....AspSerAsnV
                 :::       :::
     192 GACCCCTCAGATCCAGCAGGGTCCGA TATGTCTCCCCCCCCTTC

      99 alAspGluLeuAspThrProAspGluAlaAspSerPheG
         ::      :::       :::
     242 TGGATGATGAGCGGCGAGGCGGAAGATACTCCCCCCCCTATCCCCTA

     112 ...GluTyrThrGlyHisGlyLysSerLeuSerTrp 142
            :::   :::   :::::  :::::::::::
     292 GAAAGCTACACGCTGGAGGGAAAGTCTTTGATTGG 327

Seq_name: /tgrid_ftp1.data/1_1/na/FB_1_ME_s/US-08-832-883-1

seq_documentation_block:
; Sequence 1, Application US/08832883
; Patent No. 5840506
;  GENERAL INFORMATION:
;    APPLICANT:  Giordano, Antonio
;    TITLE OF INVENTION:  METHODS FOR THE DIAGNOSIS AND PROGNOSIS
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  116
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
;      STREET:  Suite 1800 Two Penn Center Plaza
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/832,877
;      FILING DATE:
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-13 US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4853 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..3489
US-08-832-877-1


alignment_scores:
            Quality:   73.00            Length:    112
              Ratio:  1.553              Gaps:      6
 Percent Similarity:  41.964   Percent Identity: 25.000

alignment_block:
 US-09-174-937-8 x US-08-832-877-1   ..

 Align seq 1/1  to: US-08-832-877-1  from: 1  to: 4853

      34 ProPheGluGluGlyValLeuSerProSerAlaAlaAspMetArgProGl 50
         |'|      :::|'|         |||||':::::|||..:   |||
      49 CCGGCCGCCCAGGGGTGCGCTATGCGCTCGGAGGTGACCAGTCGCCACC 98

      50 gProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysL 67
         |||||||
      99 GCCCCCG                                          105

      67 euThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIle 83
            |||          :::   |||||||||||
     106   CCTCCCCCTCCGGGGAGGGGCAGCCTCGGATGAGGAGGAG.    ..  . 144

      84 LeuSerAspAsp.....................AsnLeuAspSerPr 92
         ||||||||                        ||      ||||||
     145 ...GAGGAGGACGGGGAGGGGGAAGACGGCGCCGGTCTGCCGAGTCGCC 191

      92 e.....................AspGluIle.....AspIleAsnV 99
                               |||||:::         :::|||:
     192 CACCCCTCAGATCCAGCAGGGGTTCGACGAGCCTCTGCAGCCGCCTCAACA 241

      99 alAspGluLeuAspThrProAspGluAlaAspSerPhe........... 111
         ::|'| ||       |'|:::      ||| ':::
     242 TGGACGAGCTCGACACCCCGGACGAGGCTGACTCCTACCGCATGAGC 291

     112 ...GluTyrThrGlyHisGlyLysSerLeuSerTrp 122
            ||||||  ...|||.....||   ||
     292 CAAAGCTACACGCTGGAGGGAAAIGATCTTCATTGG 327
```

```
seq_name  /cgn2_6/ptodata/1/rna/IB_1CMB.seq.US-08-328-322-4

seq_documentation_block:
; Sequence 4, Application US/08328322
; Patent No. 5723436
;  GENERAL INFORMATION:
;    APPLICANT:  Huang, Laiqiang
;    APPLICANT:  Cyert, Martha S.
;    TITLE OF INVENTION:  Calcineurin Interacting Protein Compositions
;    TITLE OF INVENTION:  and Methods
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,322
;      FILING DATE:  24-OCT-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  P38,615
;      REFERENCE/DOCKET NUMBER:  8600-0151.10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  genomic DNA fragment containing full
;      INDIVIDUAL ISOLATE:  CNI coding sequence
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  376..3120
US-08-328-322-4


alignment_scores:
            Quality:   71.50            Length:    161
              Ratio:  1.430              Gaps:      4
 Percent Similarity:  49.505   Percent Identity: 28.713

alignment_block:
 US-09-174-937-8 x US-08-328-322-4   ..

 Align seq 1/1  to: US-08-328-322-4  from: 1  to: 3500

      36 GluGluGlyValLeuSerProSerAlaAlaAspMet.............A: 4
         |||:::   :::|||:::   |||:::| |:::::
     814 GAAGATTATCTACTTGAAAAAAGCGGTGCTTCAGTATATATGCCACCAAT 864

      48 gProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgI 65
         ||||||||||
     864 CGCTGAACCCCCTAAAGAT....    .            ..............  8

      65 leLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGly 81
                    |||   '|:|:|||||||||||      |||
```

```
      883 ..................AATAGCAATCTAAGTTTGGATGCCTATGAGCGC 915

       82 SerIleLeuSerAspAspAsnLeu...AspSerProAspGluIleAspIl 97
          :::   | ' ::::::| | ''|   :.....  |   :::   ::::: 
      916 AACTCATTGTCATCCGATAATTTGAGTAACAAGCCAGTATCAAGTGATGT 965

       97 eAsnValAspGlu.............................LeuAspT 104
          ::::   '' :::                             |||   : 
      966 TTGCCATGACGACAGTAAACTGTTGGCTATTCAAAAGACACCATTACCAT 1015

      104 hrProAspGluAlaAspSerPheGluTyrThrGlyHisGlyLysSerLeu 120
          ::   ::::::      ''|'      ...  |||||...|||||| 
     1016 CATCTAGTTGAAGAGGATCGGTACGGCAAATTTCACGGTAACTCTTTG 1065

      121 Ser 121
          ' 

     1066 TCA 1068

seq_name: /tmp/I_fgtidata/1/ina/EB_DIMB.seq.US-08-328-322-6

seq_documentation_block:
; Sequence 6, Application US/08328322
; Patent No. 5723436
;  GENERAL INFORMATION:
;    APPLICANT:  Huang, Laiqiang
;    APPLICANT:  Cyert, Martha S.
;    TITLE OF INVENTION:  Calcineurin Interacting Protein Compositions
;    TITLE OF INVENTION:  and Methods
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,322
;      FILING DATE:  24-OCT-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  P38,615
;      REFERENCE/DOCKET NUMBER:  8600-0151.10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0990
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2745 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  coding sequence of CNI
US-08-328-322-6


alignment_scores:
            Quality:   71.50            Length:    101
              Ratio:  1.430              Gaps:      4
 Percent Similarity:  49.505   Percent Identity: 28.713

alignment_block:
 US-09-174-937-8 x US-08-328-322-6   ..
```

```
 Align seq 1/1  to: US-08-328-322-6  from: 1  to: 2745

      36 GluGluGlyValLeuSerProSerAlaAlaAspMet.....  ....Ar 69
         | ::::  :::|::::   '' :::   ::::::
     439 GAAGATTATCTACTTGAAAAAAGTCGTCTTTCAGTATATATTTATAAAC 488

      48 gProGluProProAsnSerLeuAspLeuAsnAspThrHisPr ArgArgIl 87
         ' ::::::::
     489 CGCTGAACCCCCTAAAGAT.....  ....  5 7

      88 leLysLeuThrAlaProAsnIleAsnLeuAsnIleAspThrSerGlyLy 127

     508 ..................AATAGCAATCTAAGTTTGGATGCCTATGAGCGC 541

      82 SerIleLeuSerAspAspAsnLe....AspSerProAspGluIleAspIl 97
         :::   '' ' ::::   ''   ::::::   ::::   ::::::   :: 
     541 AACTCATTGTCATCCGATAATTTGAGTAACAAGCCAGTATCAAGTGATGT 590

      97 eAsnValAspGlu.....  ....  ....  ....  ....  ....Ar 130
         ::::   ' :::                                    : 
     591 TTGCCATGACGACAGTAAACTGTTGGCTATTCAAAAGACACCATTACCAT 640

     104 hrProAspGluAlaAspSerPhe....GlyHisGlyLysSerLeu 140
         ::   ::::::      ''|'      ...  ||||||...|||||| 
     641 CATCTAGTTGAAGAGGATCGGTACGGCAAATTTCACGGTAACTCTTTG 690

     121 Ser 121

     691 TCA 693

seq_name: /tmp/I_fgtidata/1/ina/I_2.MB.seq.US-08-328-322-6

seq_documentation_block:
; Sequence 2, Application US/08/211,284
; Patent No. 5985608
;  GENERAL INFORMATION:
;    APPLICANT:  Luna, Elizabeth J.
;    APPLICANT:  Pestonjamasp, Kersi N.
;    APPLICANT:  Pope, Robert M.
;    APPLICANT:  Wolfkuhle, Julie J.
;    TITLE OF INVENTION:  ACTIN-BINDING PROTEINS
;    TITLE OF INVENTION:  AND NUCLEIC ACIDS ENCODING THE SAME
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/[illegible]
;      FILING DATE:  31-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, Peter J.
;      REGISTRATION NUMBER:  32,[illegible]
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5718 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  451...5814
US-08-962-284-1


alignment_scores:
            Quality:   71.00            Length:    144
              Ratio:  1.029              Gaps:      8
 Percent Similarity:  47.917   Percent Identity:  28.472

alignment_block:
 US-09-174-937-8 x US-08-962-284-1/rev

 Align seq 1/1  to reverse of: US-08-962-284-1  from: 1  to: 6718

       4 SerCysSerArgAlaSerPheSerProSerValArgLysProProLeuIl 20
         ||||||||:::   ||||||   |||             |||   |||
    5005 TCTTGTTGGCAGCGGTCCTTCCGACCTCCTTCGTGGGCCTGGGCTTTG 4956

      20 eLeuArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGlu 36
          :::|||       :::::::|||   :::   :::||||::       :::
    4955 CATCCGTGCCACAGGTAGATGAGGGCCTT...GTTGACGTTAAGCACCAC 4909

      37 GluGlyValLeuSerProSerAlaAlaAspMet..ArgProGluProPro 52
         :::|||     ::.|||...||||||   :::  ||||||   |||
    4908 CATGGAAGTTCTGGACCTCAGGCTGCTACAGTGACAGGCCACTTCCAGCA 4859

      53 AsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThr.. 68
         ||||||| |          |||   ||              |||
    4858 AATTCCCTTCCACGGGCACCTCTCCACGCACGCAGTACAGCCGCCACTCA 4809

      69                                        AlaPro Asn 71
                                                |||||| :::
    4808 CTTTGCACATTTTCTTCTTCCTCTTCCCGCCTCTCCGAGTGCACCACCAT 4759

      72 IleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAs 88
           | :::      |||:::   || ::::::||||||   ::::||
    4758 CCCCCCCTGGAAACACTGCAGGAAACAGGGGGGTCCTTTCCCTGGAGAA 4709

      88 nLeuAsp......SerProAspGluIleAspIleAsn............. 98
         |||||||       |||      ::::::::::::
    4708 CCTGGACCTGGGGCCCCCTTTCCTCGTCCAGC TCACCGTCATCAGCGCC 4659

      99 .....                 ...ValAspGluLeuAspThrProAsp 106
                                  |||      |||      |||:::
    4658 GACGTGCCCTTCTCACTCACGGTGGAGTGCCGGCCTTG......CCAGAA 4615

     107 GluAlaAspSerPheGluTyrThrGly 115
         |||:::|||:::...  ......|||
    4614 GAAGTAGACGCACTTCTCTTTGCCGGC 4588

seq_name: /cgn2_6/ptodata/1/ima/SQ_COMM_seq US-08-741-134-5

Seq_documentation_block:
; Sequence 5, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,163
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS.
;      LENGTH:  1236 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1236
US-08-741-134-5


alignment_scores:
            Quality:   69.50            Length:    124
              Ratio:  1.178              Gaps:      5
 Percent Similarity:  47.581   Percent Identity: 24.194

alignment_block:
 US-09-174-937-8 x US-08-741-134-5  ..

 Align seq 1/1  to: US-08-741-134-5  from: 1  to: 1236

      27 GluAspValGlyMetAspIlePro[illegible]GluGlyVal.....Le 42
         ||||||         |||       : :::|:|::::|              |
    325 GAAGACGATGACGAGGACGATGAGT[illegible]CAAGAATCCGTCCTTTTGACTTT 374

      40 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 57
         |||||||:::|||                :::      ::: |:|||||||
    375 ATCTCCGGAAGCC               TAATACGAACAATCTTTTGAGT 4..

      57 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsn..... 71
         ||:::   |||      :::::::::::::::::   :::   :::
    410 TGACCATTACTCCAGAAGAAGAA[illegible]CAATTGTCACTGGTTCTTAC 459

      72 ...IleAsnLeuSer.....                  ..............  75
            ||| ::||||||
    460 GCTATCTCCTTGAGCGGTAACTATGTTAAGCATCCATTTGATACTCCAAT 51..

      75                                ..............................  75

    510 GGGAGTCGAAGGTGAAGACGAAGATGAAGACGGTGACATCTATGACAGTG 559

      76 . . .       . LeuAspGlnSerGluGlySerIleLeuSerAsp.....  89
                       |||    :::'|'   :::|||::::::||'
    560 AAGACTACGACTTGACCCCAGATGAGGATGAAACTATTGGTGACGACATT 6..

      87 AspAsnLeuAspSerProAspGl[illegible]AspIleAsnValAspGlyLeuAs 1..
         |||:::||||||:      :||:    :  :::::::  ||||||
    610 GACGACTTGGATGACGAAGAGGAAGAAGAATTCGTATTGAAGAAGTCCA 65..

    103 pThrProAspGluAlaAspSer 118
        :     || |||   |||:::
```

```
     660 AGAAGAAGATGAAGAAGATAAT 681

seq_name: /tp/1_f/pt_3ata/1/ira/US1_COMB.seq:US-08-770-379-18

seq_documentation_block:
; Sequence 18, Application US/08770379
; Patent No. 5843664
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-18


alignment_scores:
            Quality:   69.00            Length:    118
              Ratio:  1.190              Gaps:      6
 Percent Similarity:  49.153   Percent Identity: 27.119

alignment_block:
 US-09-174-937-8 x US-08-770-379-18/rev   ..

 Align Seq 1/1  to reverse of US-08-770-379-18  from 1  to 35100

      15 ArgLysProPro.....LeuIleLeuArgArgLeuLeuSerGluAs 28
         .| ::::|| ||         ::::::           ||||||:::
   33401 AGACGACCCCCCGTGAAGTTCATATTCCCACCTGCACCTCTTTCATCACT 33352

      29 pValGlyMetAspIleProPheGluGluGlyValLeuSerProSerAlaA 45
          ||:::   :||     :::|       :::|||:::::
   33351 TCCAGGATTTGGCAGGCCG......CGCGGCTATGCTGGACCCACGGTGA 33308

      45 laAspMet.............................ArgProGluPro 51
          |:|||                              |||:::|||
   33307 TAGATATGTCTGCCCCAGACGACGTCTTCGCCGAGGACACGCCATCGCCG 33258

      52 ProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuTh 68
                                                ||

   33257 CCA.....................................TAAA 33247

      68 rAlaProAsnIleAsnLeuSer.Asp...Sn.Sl.Sl.....l.l.l 94
                 ||||||||||||||          |||
   33246 CGGTCTGGATCTACGACATACCG TCAC.T.TC AC.GCC.AAC.T.AAT 33197

      94 euSerAspAspAsnLeuAspSer.lAsp.lLl..AspIl.AsnVslAsp 1
          ::::::   :::     :::   :::     :::   :::  :::
   33196 ATGACGAGGATGAGGAAGATGAA.ATGAA.AA..AAATGA.TAT.CTCAA 33147

     101 GluLeuAspThrProAsp.....GluAlaAspSerIheGlyTyrL.l 127
          ::.    :   ::::::                 :::    ::
   33146 GAGGAAGATGAGCCAGAGGGGTA.CCTGCTA.ACTTTTTCAATCTTTATA 33097

     115 yHis 116
          :
   33096 TCAC 33096

seq_name: /tp/1_f/pt_3ata/1/ira/US_COMB.seq:US-[illegible]

seq_documentation_block:
; Sequence 9, Application US/081[illegible]
; Patent No. 5821070
;  GENERAL INFORMATION:
;    APPLICANT:  LEE, WEN-HWA
;    APPLICANT:  SHAN, BEI
;    TITLE OF INVENTION:  CELLULAR GENES ENCODING
;    TITLE OF INVENTION:  RETINOBLASTOMA-ASSOCIATED PROTEINS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.[illegible] Version #1.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/139,937
;      FILING DATE:  20-OCT-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/[illegible]
;      FILING DATE:  20-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN
;      REGISTRATION NUMBER:  31,[illegible]
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  688 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-139-937-9


alignment_scores:
            Quality:   67.50            Length:    [illegible]
              Ratio:  1.274              Gaps:      [illegible]
 Percent Similarity:  50.476   Percent Identity: [illegible]

alignment_block:
 US-09-174-937-8 x US-08-139-937[illegible]   ..
```

```
 Align seq 1/1  to  US-08-174-937-8  from  1  to  688

      42 ProSerAlaAlaAspMetArgProGluProPro ..... 52
         ||||||||||      ::::::   :::   |||
     142 CCTAGTGCAAAGAAAGTGAAGAAACGGAAICCTTGGTCAGATGATGAAIC 191

      53 ....AsnSerLeuAspLeuAsnAspThrHisPro..... 62
             ::::::   ||||||::::::|||:::| :
     192 CAACTCAGAAAGTGATTTGGAAGAAACAGAACCTGTGGTTATTCCAAGAG 241

      63 ...........ArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
                   |||||                ||:::  :::::::::
     242 ATTCTTTGCTTAGGAGAGCAGCAGCCGAAAGACCTAAATACACATTTAAT 291

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ::::::::::::::|||   :::      :::|| ||||::   |||:::
     292 TTTTCAGAATAAGATGATGATGTTGATTATGATGATTACAATAA 341

      92 uAspGluIleAspIleAsnVal ..... 99
          |||      .........|||
     342 TGATTTAGAGGAATTGAAAGTTAAAGCATCTCTTATAACAAATGATGGGG 391

     100 ..AspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThr... 114
           ||||:::      ||||      ||||::   |||||||||
     392 AAGATGAATTTGTTCCTTCAGATGGGTTAGATAAAGATGAATATACATTT 441

     115 GlyHisGlyLysSer 119
         :::   ||||||
     442 TCACCAGGCAAATCA 456

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMP.seq:PCT-US08-00174-7

seq_documentation_block:
; Sequence 7, Application US/08591079
; Patent No. 5972899
;  GENERAL INFORMATION
;    APPLICANT:  Zychlinksky, Arturo
;    APPLICANT:  Chen, Yajing
;    TITLE OF INVENTION:  Apoptosis Induced by Shigella IpaB
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,079
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Livnat, Shmuel
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  15661-20017.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0764
;      TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3622 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ORIGINAL SOURCE:
;      ORGANISM:  Salmonella typhimurium
;      STRAIN:  sibB
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  575..2356
;      OTHER INFORMATION:  /product= "SipB"
;      OTHER INFORMATION:  /gene= "sipB"
US-08-591-079-7


alignment_scores:
            Quality:   67.50              Length:    113
              Ratio:  1.125                Gaps:      4
 Percent Similarity:  53.097   Percent Identity: 20.354

alignment_block:
 US-09-174-937-8 x US-08-591-079-7   ..

 Align seq 1/1  to  US-08-591-079-7  from  1  to  3622

      11 SerProSerValArgLysProProLeuIleLeuArgArgLeuLeuSerIl 27
         |||   .....:||||||||||
    3116 TCGTTAAATATAAGAAAACCG..                ..............  3160

      27 uAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSerA 44
                  :::|||:::::::::   ::::|:
    3137 ..............GTACCGATGCGACGAAAAATCTTAATGATGCGAC  3177

      44 laAlaAspMetArgProGluPro...ProAsnSerLeuAspLeuAsnAsp 59
                :::::   |||||||||   ''':::::::::   :::
    3171 GCTTAAATCTAATGCCGGAACCAGCCCACCGAAAGTCTCGGTATTAAAT 3221

      60 ThrHisProArgArgIleLysLe[illegible]laProAsnIleAsnLeuSerIe 76
         |||      :::::::::::::||::::          ::::::
    3221 ACAGIAATAAACAAAGICICCCIGAACATCTATAICTTGTCGAAACGTCT 3271

      76 uAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerProA 94
         :::  :::   |||:  :::|||      ::::::|||
    3271 TGAGTCTGTCGAATCCGATATTCGTCTTGAGCAGAATTAC.......... 3311

      93 spGluIleAspIleAsnValAspGluLeuAspThrProAspGluAlaAsp 109
                 :::::::::   ::::::|||:::   ::::::|||| |
    3311  ..         ATGGATATTACCCGAATCGATAGCGCGCAAGATGCAGAT 3354

     110  ...         SerPheGluTyrThrGlyHisGlyLys 118
                      :::::|||      '|||||'|||::
    3350 GACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGG 3388

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMP.seq:PCT-US93-11310-9

seq_documentation_block:
; Sequence 9, Application PC/TUS9311310
;  GENERAL INFORMATION:
;    APPLICANT:  BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
;    TITLE OF INVENTION:  CELLULAR GENES ENCODING
;    TITLE OF INVENTION:  RETINOBLASTOMA-ASSOCIATED PROTEINS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11310
;      FILING DATE:  19-NOV-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN
```

```
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  FP-CI 4750
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-535-9001
;       TELEFAX:  619-535-8949
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  688 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
PCT-US93-11310-9


alignment_scores:
            Quality:   67.50            Length:    105
              Ratio:  1.274              Gaps:      4
 Percent Similarity:  50.476   Percent Identity: 25.714

alignment_block:
 US-09-174-937-8 x PCT-US93-11310-9   ..

 Align seq 1/1  to:  PCT-US93-11310-9  from:  1  to:  688

      42 ProSerAlaAlaAspMetArgProGluProPro..................... 52
         ||||||||.      ::::::   :::    |
     142 CCTAGTGCAAAGAAAGTGAAGAAACGGAATCCTTGGTCAGATGATGAATC 191

      53 ....AsnSerLeuAspLeuAsnAspThrHisPro..................... 62
             ::::::   :|:|:|:::::::|||:::|
     192 CAACTCAGAAACTGATTTGGAAGAAATAGAACCTGTGGTTATTCCAAGAG 241

      63 ............ArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
                     |:|:|          |:::   ::::
     242 ATTCTTTGCTTAGGAGAGCAGCAGCCGAAAGACTTAAATACACATTTAAT 291

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ::::::::::::::|  :::      ::: |  |::::   ||||::
     292 TTGTCAGAAGAAGAGGATGATGATGCTGATGATGATGATGACAATAA 341

      92 pAspGluIleAspIleAsnVal                          . 99
         | .      ::::::::: |
     342 TGATTTAGAGGAATTGAAAGTTAAAGCATCTCCCATAACAAATGATGGGG 391

     100 ..AspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThr... 114
           ' ' |:::                ::   |||| |'|
     392 AAGATGAATTTGTTCCTTCAGATGGGTTAGATAAAGATGAAATATACATTT 441

     115 GlyHisGlyLysSer 119
         :::   ' ||
     442 TCACCAGGCAAATCA 456

seq_name.  /tp/2_f/pt/data/1/fsa/DA_COMB.seq.US-07-702-771-2

seq_documentation_block:
; Sequence 2, Application US/07702771
; Patent No. 5610276
;  GENERAL INFORMATION:
;     APPLICANT:  Wong, Gail L.
;     APPLICANT:  McCormick, Frank
;     TITLE OF INVENTION:  Cloning and Purification of P62
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Cetus Corporation
;       STREET:  1400 Fifty-Third Street
;       CITY:  Emeryville
;       STATE:  California
;       ZIP:  94608-2997
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.4
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/7..2,771
;       FILING DATE:  17-MAY-1991
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Giotta, Gregory J.
;       REGISTRATION NUMBER:  32,...8
;       REFERENCE/DOCKET NUMBER:  ...4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-420-3300
;       TELEFAX:  415-658-5239
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1485 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  N
;     ANTI-SENSE:  N
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
US-07-702-771-2


alignment_scores:
            Quality:   67.00            Length:    ...
              Ratio:  0.971              Gaps:      ...
 Percent Similarity:  50.000   Percent Identity:  ...

alignment_block:
 US-09-174-937-8 x US-07-702-771-2   ..

 Align seq 1/1  to:  US-07-702-771-2  from:  1  to:  ...

       8 AlaSerPheSerProSerValAr...ysPr...r...euIle...Ar...Ar...  ..
                                 |  ::::::::::          :::
      23 GCCCGGGCCTCGCCCGCCACGCA...C...C...G...T...C...T...C...  ..

      24 uLeuSerGluAspValGlyMetAs...Ile...r...PheGl...G...G...V...  ..
         :::    ' :::   :::                          :::
      73 CACGGGTCCCGACGCGACAGTGG...T...T...A...C...A...T...T...C...  ...

      41 erProSerAla...AlaAspMetAr...r...luProProAsn.....  ..
         ' ' ' ' ' |  :::::: ::
     123 CCCCCTCGGCCACAGCCTCGGTCA...A...T...A...A...A...A...T...C...  ...

      54 ..........................erP...AspLeuAs...As...Ile...  ..
                                                 ::
     173 CCCGAACTCATGGCCGAGAAGGA...T...C...A...C...G...T...A...T...  ...

      62 pArgArgIleLysLeuThrAlaP...AsnI...AsnLe...Ser...Asp...  ..
         ::::::   ' ' '          ' ::::::::          :::
     223 CATCCAGTTG...GTGACGGCAGAA...A...A...A...A...A...A...T...  ...

      79 erGluGly.....................rIleL...AspAspAsn...  ..
         ''::::                      ::::::::
     271 CAAAAAAGGATGATGAGGAAI...I...A...T...I...A...A...  ...

      89 LeuAspSerProAspGluIleAs...IleAsnValAsp...LeAs...  ..
         :::          :::::::::                 :::
     320 ATGAAACTGAAAGAGCGAGTGCT...A...C...T...T...A...A...T.....  ...

     105 pAspGluAlaAspSerPheGlu...yTh...l,...His...Gl,...L...S...  ...
                    ::::'|::::::::::          '  '          :::
     362 ..........CCCAAGTTCAATT...T...T.....GG...AA...A...T...  ...

     122 rpGlnGlyGlnSer 126
         | ' '|:::::
     396 CACAAGGGAATACA 409

seq_name.  /tp/2_f/pt/data/1/fsa/PA_2_MB.seq.US-...-...-...-...
```

```
seq_documentation_block:
; Sequence 2, Application US/08438883
; Patent No. 5731427
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, G.
;    APPLICANT:  McCormick, F.
;    TITLE OF INVENTION:  Purification and Cloning of p62
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/438,883
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/702,771
;      FILING DATE:  17-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pochopien, Donald J.
;      REGISTRATION NUMBER:  32,167
;      REFERENCE/DOCKET NUMBER:  27527/33275
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1485 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-438-883-2


alignment_scores:
            Quality:   67.00              Length:    138
              Ratio:  0.971                Gaps:      6
 Percent Similarity:  50.000   Percent Identity: 25.362

alignment_block:
 US-09-174-937-8 x US-08-438-883-2   ..

 Align seq 1/1  to: US-08-438-883-2  from: 1  to: 1485

       8 AlaSerPheSerProSerValArgLysProPr LeuIleLeuArgArgLe 24
         |||      ||||||:::::::::::   ||;||. ||:::
      25 GCCCGGGGCCCGGCCCGCCACGCAGCCGCCACCC-CGCTGCCGCCCTCGGC 72

      24 cLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValLeuS 41
           :::   |||:::   :::       |'|           :::|||
      73 CACGGGTCCCGACGCGACAGTGGGGGGGCCAGCGCCGACCCCGCTGCTGC 122

      41 erProSerAla...AlaAspMetArgProGluProProAsn..... 53
         |||||||||   |||:::::::::   ||||.|   |||
     123 GCCCCTCGGCCACAGCCTCGGTCAAGATGGAGTTAGAGAACAAGTACCTG 172

      54 .....SerLeuAspLeuAsnAspThrHisPr 62
                ||||||||    :::   ||||||

173 CCCGAACTCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACTCACGC 222

 62 cArgArgIleLysLeuThrAlaP:cAsnIleAsnLeuSerIleAspGlnS 7+
      ::::::   ||||||||||          ::::::::::       ::::
223 CATGCAGCTG...CTGACGGCAGAAATTGAGAAGATTCAGAAAGGAGACT 269

 79 erGluGly... ........ .       SerIleLeuSerAspAspAsn 94
    ||:::                       .......:::||        |::
270 CAAAAAAGGATGATGAGGAGAATTACTTGGATTTATTTTCTCATAAGAAC 319

 89 LeuAspSerProAspGluIleAspIleAsnValAspGluLeuAspThrPr 1 5
    :          ::::::          '||  '||   :::
320 ATGAAACTGAAAGAGCGAGTGCTGATAGCTGTCAAGCAGTAT..... 361

105 cAspGluAlaAspSerPheGluTyrThrGlyHisGlyLysSerLeuSerI 122
                 ...|||.....:::::       ||||||   ||:::
362           .CCCAAGTTCAATTTTGTG      GGGAAGATTCTTGGAC 395

122 rpGlaGlyGlnSer 126
     ||||||:::::
396 CACAAGGGAATACA 409
```

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 08:54:41 ; Search time 5888.89 Seconds
(without alignments)
-194.901 Million cell updates/sec

Title: US-09-174-937-9
Perfect score: 378
Sequence: 1 atgctgaaagagctgtagtcg taagttggcaaggccaaagt 378

Scoring table: IDENTITY_NUC

Searched: 921193 seqs, 15181192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold : 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score histogram.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	282.4	74.7	5654	9	AB002365	AB002365 Human mRN
2	281	61.1	169798	44	AJ000581	AJ000581 Homo sapi
3	100.4	26.6	179201	41	AC010181	AC010181 Homo sapi
4	68.8	18.2	169798	44	AJ000581	AJ000581 Homo sapi
5	37	9.8	446	11	AA478...	AA478... AA740_1 Zeb
6	35.2	9.3	117897	35	AP001454	AP001454 Homo sapi
7	35	9.3	172866	41	AC011441	AC011441 Homo sapi
8	34.4	9.1	113950	8	AIAT_4647	AC 4647 Arabidops
9	34.4	9.1	1740	12	MMSTIF	X... 2 Mus musculus
10	34.4	9.1	1858	12	MMIFANS1	X... 2 Mus musculus
11	34.4	9.1	130027	41	AC012481	AC 2481 Plasmodiu
12	34.4	9.1	300729	42	AC010424	AC 0424 Homo sapi
13	34.2	9.0	10066	2	AF...	... ... Staphylococcus
14	34.2	9.0	92611	8	AI1122	AC... Arabidops
15	34.2	9.0	172031	44	AC012069	AC 2069 Homo sapi
16	34	9.0	1366	5	AB 28473	AB 8473 Sequence
17	34	9.0	2382	11	HSPIF...	HSPIF... H.sapien
18	34	9.0	166826	44	AC012...	AC 2... Homo sapi
19	33.6	8.9	40548	7	AB 1...	AB 1... Arabidops
20	33.6	8.9	132400	33	AC... 41	AC 41 Homo sapi
21	33.6	8.9	5272	35	AF... 8	AF 8 Plasmodiu
22	33.4	8.8	1240	4	BA...	... Bacillus alc
23	33.4	8.8	4720	34	...	... Hemorhabdi
24	33.4	8.8	150985	43	AC...	AC ... Homo sapi
25	33.2	8.8	5224	12	AF...	AF ... Homo sapi
26	33.2	8.8	4261	11	HTMSLCKIN3	M... ... Human gluco
27	33.2	8.8	595	26	HTMDT79H3	L... ... Human STS D
28	33.2	8.8	167228	47	AC...	AC ... Homo sapi
29	33.2	8.8	153201	46	AC...	AC ... Homo sapi
30	33.2	8.8	127764	43	AC...	AC ... Homo sapi
31	33	8.7	35769	8	SPAC...	AL ... S.pombe c
32	32.8	8.7	191692	32	DNA...	AC ... Homo sapi
33	32.8	8.7	189406	33	DNA...	AC ... Homo sapi
34	32.8	8.7	149643	42	AC...	AC ... Homo sapi
35	32.6	8.6	85572	33	...	... Homo sapi
36	32.6	8.6	95025	38	HS... D12	AC ... Homo sapi
37	32.6	8.6	187639	42	AC...	AC ... Homo sapi
38	32.6	8.6	190248	45	AC...	AC ... Homo sapi
39	32.4	8.6	29142	11	HSF...	AF ... Human DNA s
40	32.4	8.6	45199	11	HSP...	AC ... Human DNA
41	32.4	8.6	206240	16	AC...	AC ... Melanoplu
42	32.4	8.6	157056	43	AC...	AC ... Homo sapi
43	32.4	8.6	178834	44	AC...	AC ... Homo sapi
44	32.4	8.6	176956	44	AC...	AC ... Homo sapi
45	32.2	8.5	3831	2	AF...	AF ... Mycoplasm

ALIGNMENTS

RESULT 1
AB002365
LOCUS AB002365 5654 bp mRNA PRI ...-FEB-1999
DEFINITION Human mRNA for KIAA... gene, partial cds.
ACCESSION AB002365
VERSION AB002365.1 GI:2224674
KEYWORDS KIAA0367.
SOURCE Homo sapiens male brain cDNA to mRNA ... clone:IMAGp...II SK plus clone:HH0141.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1, 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA0367"
     CDS             <1..2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20822.1"
                     /db_xref="GI:2204675"
                     /translation="ALEDGDPHLSTENPALVPDALLASDTCLDISEAAFDHSFSDASG
                     LNTSTGTIDDMSKLILSEGHPEIPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDLL
                     PEDPHFLYGDFPLEEDSIFLGLAPYTFPFTLSYLTEPAQSAETIEEAGSPEDESLGC
                     RAAEIVLSALPDPRSEGNQAETKNPLPGSQLAVLHIPEDPESVYLPVGAGSNILSPSN
                     VDWEVETDNSDLFAGSIIGPENGASKEIPELEEEKTIPTKEPEQIKSEYKEERGTEKN
                     EDRHALHMDYILVNPEENSHSKPETCEERESIAELELYVGSKETGLQGTQLASFPDTC
                     QPASLNERKGLSAEKMSSKGDTPSSFESPAQDQSWMFLGHSEVGDPSLLDAPDSGPGWS
                     GKTVEPPSPIGLGRGPQLQIPEEMKPIESLAIEEASGPVSQSQKSKSPGRAGPDAVTH
                     DSEWEMLSPQPVQKNMISQILMEEEIEFLELGTRISRPNGLLSEDVGMDIPFEEGVLS
                     PSAADMRPEPPNSLDLNDTHFRIKLIAPNINLSLDQSEGSILSDDNLDSPDEIDINV
                     DELDTPDEADSFEYTGHDPTANKDSGQESESIPEYTAEEEREDNPLWPTVVIGEQEQP
                     IDMFVIEPYPPVISHGGYYGTILNAIIVFAACFLPDSSPADYHYVMENLFLYVISTLE
                     LMVAEDYMIVYLNGATPKKEDAPAPLDEEMLPDD"
BASE COUNT      1664 a    1188 c    1198 g    1604 t
ORIGIN


  Query Match           74.7%;  Score 283.4;  DB 9;  Length 5654;
  Best Local Similarity   99.6%;  Pred. No. 1.3e-71;
  Matches  283;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      68 gaatactatcaaaaaatgtaaggaatggaatatc cctttgaagaggggcgtgctgagtccca 127
           ||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db    1469 GAATACTGTCAGAGGATGTAGGAATGGAATATCCCCTTTGAAGAGGGCGTGCTGAGTCCCA 1528

Qy     128 gtgctgtggagcatgtggatctaaatctcttatgtggatcttaatgacactcatcctc 187
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    1529 GTGCTGCAGACATGAGGCCTGAACCTCCTAAATCTCTGGATCTTAATGACACTCATCCTC 1588

Qy     188 ggaaatcaagctcacagcccaaaatatcaaatttctctggaccaaagtgaaggatcta 247
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1589 GGAGAATCAAGCTCACAGCCCCAAAATATCAAATTTCTCTGGACCAAAGTGAAGGATCTA 1648

Qy     248 ttctcttgatgataaattttaaaatccaatcaaattgacatcaatgtggatgaacttg 307
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1649 TTCTCTTGATGATAAATTGGACAGTCCAGATAAAATTGACATCAATGTGGATGAACTTG 1708

Qy     308 atagccccgatgaaagcagattcttttgaata actggccatgt 351
           |||||||||||||||||||||||||||||| ||||||||||||
Db    1709 ATAGCCCCGATGAAGCAGATTCTTTTGAATA ACTGGCCATGT 1752
```

```
RESULT   2
AC009556
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 17 clone RP11-57321 map 17, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-57321
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barry,A., Beckerle,M., Berry,A., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,F., Doran,K., Dowes,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,J., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liu,C., Locke,K., MacDonald,P.,
            Marquis,N., McEwan,P., McCurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Myrabalen-Kyaj,T.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Parkins,R.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Ruy,A., Savery,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vil,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yo,W.I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6479157.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: LL672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13: M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136295 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 7.2 in Q20 bases; agarose-fp
                Quality coverage: 8.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *    1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *    2037     3371: contig of 1335 bp in length
```

```
                *                          gap of unknown length
                *     3372     4972: contig of 1601 bp in length
                *                          gap of unknown length
                *     4973     5981: contig of 1009 bp in length
                *                          gap of unknown length
                *     5982     8008: contig of 2027 bp in length
                *                          gap of unknown length
                *     8009     9286: contig of 1278 bp in length
                *                          gap of unknown length
                *     9287    11243: contig of 1957 bp in length
                *                          gap of unknown length
                *    11244    14312: contig of 3069 bp in length
                *                          gap of unknown length
                *    14313    18423: contig of 4111 bp in length
                *                          gap of unknown length
                *    18424    23065: contig of 4642 bp in length
                *                          gap of unknown length
                *    23066    30558: contig of 7493 bp in length
                *                          gap of unknown length
                *    30559    37537: contig of 6979 bp in length
                *                          gap of unknown length
                *    37538    44831: contig of 7294 bp in length
                *                          gap of unknown length
                *    44832    50944: contig of 6113 bp in length
                *                          gap of unknown length
                *    50945    58263: contig of 7319 bp in length
                *                          gap of unknown length
                *    58264    65216: contig of 6953 bp in length
                *                          gap of unknown length
                *    65217    72310: contig of 7094 bp in length
                *                          gap of unknown length
                *    72311    78799: contig of 6489 bp in length
                *                          gap of unknown length
                *    78800    87607: contig of 8808 bp in length
                *                          gap of unknown length
                *    87608    98337: contig of 10730 bp in length
                *                          gap of unknown length
                *    98338   112198: contig of 13861 bp in length
                *                          gap of unknown length
                *   112199   126330: contig of 14132 bp in length
                *                          gap of unknown length
                *   126331   140617: contig of 14287 bp in length
                *                          gap of unknown length
                *   140618   153843: contig of 13226 bp in length
                *                          gap of unknown length
                *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a   34914 c   34487 g   49065 t      2 others
ORIGIN


  Query Match             61.1%;  Score 231;  DB 44;  Length 169798;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-56;
  Matches  231;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     148 gaacctcctaattctctggatcttaatgacactcatcctcggagaatcaagctcacagcc 207
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119825 GAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCC 119884

Qy     208 ccaaatatcaatctttctctggaccaaagtgaaggatctattctctctgatgataacttg 267
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119885 CCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTG 119944

Qy     268 gacagtccagatgaaattgacatcaatgtggatgaacttgatacccccgatgaagcagat 327
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119945 GACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCCGATGAAGCAGAT 120004
```

```
Qy     328 tcttttgagtacactggccatg                                      348
           ||||||||||||||||||||||                                       
Db  120005 TCTTTTGAGTACACTGGCCATG                                      120025


RESULT   3
AC011488/c
LOCUS       AC011488   179301 bp    DNA             HTG        [illegible]-1999
DEFINITION  Homo sapiens chromosome 19 clone [illegible], *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact lengths of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1     1007: contig of 1007 bp in length
            *                      gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *                      gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *                      gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *                      gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *                      gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *                      gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *                      gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *                      gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *                      gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *                      gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *                      gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *                      gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *                      gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *                      gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *                      gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *                      gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *                      gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *                      gap of unknown length
            *    13948    15016: contig of 1069 bp in length
```

```
*                    gap of unknown length
*       15017    15614: contig of 598 bp in length
*                    gap of unknown length
*       15615    16687: contig of 1073 bp in length
*                    gap of unknown length
*       16688    18152: contig of 1465 bp in length
*                    gap of unknown length
*       18153    19221: contig of 1069 bp in length
*                    gap of unknown length
*       19222    20477: contig of 1256 bp in length
*                    gap of unknown length
*       20478    21388: contig of 911 bp in length
*                    gap of unknown length
*       21389    22058: contig of 670 bp in length
*                    gap of unknown length
*       22059    23154: contig of 1096 bp in length
*                    gap of unknown length
*       23155    23925: contig of 771 bp in length
*                    gap of unknown length
*       23926    24291: contig of 366 bp in length
*                    gap of unknown length
*       24292    24907: contig of 616 bp in length
*                    gap of unknown length
*       24908    25839: contig of 932 bp in length
*                    gap of unknown length
*       25840    27116: contig of 1277 bp in length
*                    gap of unknown length
*       27117    27904: contig of 788 bp in length
*                    gap of unknown length
*       27905    28519: contig of 615 bp in length
*                    gap of unknown length
*       28520    28930: contig of 411 bp in length
*                    gap of unknown length
*       28931    30384: contig of 1454 bp in length
*                    gap of unknown length
*       30385    31245: contig of 861 bp in length
*                    gap of unknown length
*       31246    32148: contig of 903 bp in length
*                    gap of unknown length
*       32149    33058: contig of 910 bp in length
*                    gap of unknown length
*       33059    34020: contig of 962 bp in length
*                    gap of unknown length
*       34021    34782: contig of 762 bp in length
*                    gap of unknown length
*       34783    36041: contig of 1259 bp in length
*                    gap of unknown length
*       36042    37033: contig of 992 bp in length
*                    gap of unknown length
*       37034    37565: contig of 532 bp in length
*                    gap of unknown length
*       37566    38344: contig of 779 bp in length
*                    gap of unknown length
*       38345    39359: contig of 1015 bp in length
*                    gap of unknown length
*       39360    41242: contig of 1883 bp in length
*                    gap of unknown length
*       41243    41839: contig of 597 bp in length
*                    gap of unknown length
*       41840    42836: contig of 997 bp in length
*                    gap of unknown length
*       42837    43913: contig of 1077 bp in length
*                    gap of unknown length
*       43914    55087: contig of 11174 bp in length
*                    gap of unknown length
*       55088    81867: contig of 26780 bp in length
*                    gap of unknown length
*       81868   112600: contig of 30733 bp in length
*                    gap of unknown length
*      112601   179301: contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42598 t     85 others
ORIGIN


  Query Match             26.6%;  Score 100.4;  DB 41;  Length 179301;
  Best Local Similarity   64.8%;  Pred. No. 8.4e-19;
  Matches  149;  Conservative    0;  Mismatches   81;  Indels     :  Gaps    0;

Qy    146 ctgaacctcctaatttctgga[illegible] 205
          | |  ||||| ||  | ||  || | || |   | |||| |  | | | | |
Db 124306 CAGCTCCTCCCAACACCCTAAA[illegible] 124247

Qy    206 cccaaatatcaatctttctctg[illegible] 265
          ||||| | |||||  |||||||  | | |||| | | |  | | | | |
Db 124246 CCCCACAGATCAACCATTTCTCT[illegible] 124187

Qy    266 tggacagtccagatgaaattga[illegible] 325
          |||| |  || |||||  | || |  || |||||||  | | | | |
Db 124186 TGGATACCCTTGATGACCTGGA[illegible] 124127

Qy    326 attcttttgagtacactggccatggtaagtcactaagttgccaaaggccaa 375
          | ||  | |||| |   ||  | || ||  ||  |  ||, |||  ||
Db 124126 ACTTGTTGGATTCCTTGGGAA[illegible] 124 77


RESULT   4
AC009556/c
LOCUS       AC009556   169798 bp    DNA              HTG       8-N V-1999
DEFINITION  Homo sapiens chromosome 19, clone RP11-57C21, [illegible] DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 19, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,P., Bennett,A., Brown,A.,
            Castle,A., Cerny,J., Christopher,M., Collins,S., Gillmore,A.,
            Cooke,P., DeArellano,K., Depayre,F., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrester,W.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,L., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liew,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., Mesirov,A., McKernan,K., McPherson,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Parkin,R.,
            Peterson,K., Pollara,V., Riley,R., Roberts,L., Ryan,A., Sekhon,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,G., Vo,A., Warner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yao,W.T. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) [illegible] Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced [illegible].
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            ------------- Project Information
```

```
    Center project name: L1672
    Center clone name: 57_C_21
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-primer-amersham; 5% of reads
    Chemistry: Dye-terminator Big Dye; 95% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 136395 bases at least Q40
    Consensus quality: 156543 bases at least Q30
    Consensus quality: 165534 bases at least Q20
    Insert size: 165000; agarose-fp
    Insert size: 169798; sum-of-contigs
    Quality coverage: 4.2 in Q20 bases; agarose-fp
    Quality coverage: 4.1 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1016: contig of 1016 bp in length
*                   gap of unknown length
*     1017     2036: contig of 1020 bp in length
*                   gap of unknown length
*     2037     3371: contig of 1335 bp in length
*                   gap of unknown length
*     3372     4972: contig of 1601 bp in length
*                   gap of unknown length
*     4973     5981: contig of 1009 bp in length
*                   gap of unknown length
*     5982     8008: contig of 2027 bp in length
*                   gap of unknown length
*     8009     9286: contig of 1278 bp in length
*                   gap of unknown length
*     9287    11243: contig of 1957 bp in length
*                   gap of unknown length
*    11244    14312: contig of 3069 bp in length
*                   gap of unknown length
*    14313    18423: contig of 4111 bp in length
*                   gap of unknown length
*    18424    23065: contig of 4642 bp in length
*                   gap of unknown length
*    23066    30558: contig of 7493 bp in length
*                   gap of unknown length
*    30559    37537: contig of 6979 bp in length
*                   gap of unknown length
*    37538    44831: contig of 7294 bp in length
*                   gap of unknown length
*    44832    50944: contig of 6113 bp in length
*                   gap of unknown length
*    50945    58263: contig of 7319 bp in length
*                   gap of unknown length
*    59264    65216: contig of 5953 bp in length
*                   gap of unknown length
*    65217    72310: contig of 7094 bp in length
*                   gap of unknown length
*    72311    78799: contig of 6489 bp in length
*                   gap of unknown length
*    78800    87607: contig of 8808 bp in length
*                   gap of unknown length
*    87608    98337: contig of 10730 bp in length
*                   gap of unknown length
*    98338   112198: contig of 13861 bp in length
*                   gap of unknown length
*   112199   126330: contig of 14132 bp in length
*                   gap of unknown length
*   126331   140617: contig of 14287 bp in length
*                   gap of unknown length
*   140618   153843: contig of 13226 bp in length
*                   gap of unknown length
*   153944   169798: contig of 15955 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  44061 t      2 others
ORIGIN


  Query Match           18.2%;  Score 68.8;  DB 44;  Length 169798;
  Best Local Similarity   96.4%;  Pred. No. 1.2e-14;
  Matches   76;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 atgctgaagagctgtagtcgtgcatgcttctttacc tcgttttgaaaaaccttcgtttgtaatc 60
           ||||||||||||||||||| ||| | | | | | | | | | | | | | | | | | | | | |
Db   97933 ATGCTGAAGAGCTGTAGTCGTGCTGCA.TCTTCTTACCTCGTTTTGAAAAACCTTCGTTTGTAATC 97874

Qy      61 ctcagaagactactgtcagaggatgtag 88
           ||||||||  ||  | | | | | | |
Db   97873 CTCAGAAGGTAGGTGGGTCCTGGTGTAG 97846


RESULT   5
G45875
LOCUS       G45875        446 bp    DNA             STS       29-MAR-1999
DEFINITION  Z4740_1 Zebrafish AB Danio rerio STS genetic marker Z4740, 5',
            sequence tagged site.
ACCESSION   G45875
VERSION     G45875.1  GI:4492166
KEYWORDS    STS.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.
            and Fishman,M.C.
  TITLE     A genetic linkage map of the zebrafish with 2.1 microsatellite
            markers
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Mark C. Fishman
            Cardiovascular Research Center
            Massachusetts General Hospital
            Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
            Fax: 6177265806
            Email: fishman@mgh.cvrc.harvard.edu
            http://zebrafish.mgh.harvard.edu
            Primer A: CAAACAGCAGCAGCAGAGAC
            Primer B: TGTGGTCAAACACCCAGAAA
            STS size: 224
            PCR Profile:
                Presoak:          94 degrees C for 5.. minutes
                Denaturation:     94 degrees C for 1.. minute
                Annealing:        58 degrees C for 1.. minute
                Polymerization:   72 degrees C for 1.5 minute
                PCR Cycles:       25
                Thermal Cycler: MJ Research PTC-1..
            Protocol:
                Template:         10 ng
                Primer:           each 375 nM
                dNTPs:            each 200 uM
                Taq Polymerase:   0.04 units/ul
                Total Vol:        10 ul

            Buffer:
                              MgCl2:              1.5 mM
                              KCl:                50 mM
                              Tris-HCl:           10 mM
```

```
                     pH.             8.3.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="24740"
                     /clone_lib="Zebrafish AB"
                     /sex="F"
                     /dev_stage="Adult"
                     /lab_host="DH5alphaF'IQ"
                     /note="Vector: m13MP19 with added BstXI site; V-type
                     Phage; Genomic DNA  from a single adult Zebrafish of AB
                     strain was digested with  AluI,  Cac8I, HaeIII, NlaVI, or
                     RsaI.  Fragments in the range of 250-500 bp were  gel
                     purified and a BstX1 linker was added.  The fragments were
                     cloned into a  modified M13mp19 vector and transformed
                     into E. Coli DH5alpha.  Microsatelllite sequences were
                     screened with labeled d(CA)15 and d(GT)15 oligonucleotide
                     probes."
     STS             4. .227
     primer_bind     4. .23
     primer_bind     complement(209. .227)
BASE COUNT       82 a     56 c    136 g    165 t      7 others
ORIGIN


  Query Match             9.8%;  Score 37;  DB 13;  Length 446;
  Best Local Similarity   52.2%;  Pred. No. 1.4;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    181 catcctcgaggaatcaagctcaacagccccaaatatcaatctttctctggaccaaagtgaa 240
          || |   | |||| || ||  |||||| | || |     ||| |  | ||  ||   |||
Db     11 CAGCAGCAGAGACTCCAGACCACAGCACAAACTGGACTTCTATAACAGGTACATGATGAT 70

Qy    241 agatctattctctctgatgataacttggacaatccagatgaaattgacatcaatgtggat 300
          |   | |   |||||||  |   |||  |  |||||  |||    |||  ||
Db     71 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAC 130

Qy    301 gaacttgataccccccgatgaagcagattctttgagt 337
          ||   ||||      ||||| |  |||  |  |||||
Db    131 GACGATGATGATGATGATGATGATGATGATGAGT 167


RESULT   6
AF130359
LOCUS       AF130359   117937 bp    DNA             HTG       29-MAR-1999
DEFINITION  Homo sapiens chromosome 21 clone PAC K15697 map 21q21, ***
            SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION   AF130359
VERSION     AF130359.1  GI:4530581
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117937)
  AUTHORS   Polley,A., Hildmann,T., Dagand,E., Nordsiek,G., Drescher,B.,
            Weber,J., Schattevoy,R., Yaspo,M.L. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     1-4010 contig of 4010bp 4011-4111 gap of unknown size 4011-19649
            contig of 15539bp 19649-19650 gap of unknown size 19650-24324
            contig of 14675bp 24324-24325 gap of unknown size 24325-38033
            contig of 3709bp 38033-38034 gap of unknown size 38034-45456 contig
            of 7423bp 45456-45457 gap of unknown size 45457-48780 contig of
            3324bp 48781-48781 gap of unknown size 48781-51549 contig of 2769bp
            51549-51550 gap of unknown size 51550-53506 contig of 1957bp
            53506-53507 gap of unknown size 53507-58315 contig of 4809bp
            58315-58316 gap of unknown size 58316-61793 contig of  3478bp
            61793-61794 gap of unknown size 61794-77340 contig of 15547bp
            77340-77341 gap of unknown size 77341-81096 contig of 3756bp
            81096-81097 gap of unknown size 81097-84138 contig of 3042bp
            84138-84139 gap of unknown size 84139-87627 contig of 3489bp
            87627-87628 gap of unknown size 87628-94886 contig of 7259bp
            94886-94887 gap of unknown size 94887-100661 contig of 5775bp
            100661-100662 gap of unknown size 100662-103246 contig of 2585bp
            103246-103247 gap of unknown size 103247-112036 contig of 8790bp
            112036-112037 gap of unknown size 112037-117283 contig of 5247bp
            117283-117284 gap of unknown size 117284-117939 contig of 656bp.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved
FEATURES             Location/Qualifiers
     source          1. .117937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21"
                     /clone="PAC K15697"
     repeat_region   complement(1. .137)
                     /rpt_family="L1M1"
                     /evidence=not_experimental
     repeat_region   complement(138. .450)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     misc_feature    192. .291
                     /note="GC score = 4.00 (100bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     exon            333. .472
                     /note="MZEF, score = 99.6%"
                     /evidence=not_experimental
     repeat_region   complement(451. .3781)
                     /rpt_family="L1M1"
                     /evidence=not_experimental
     exon            complement(957. .1029)
                     /note="GRAIL, score = 48.000%, comment = marginal"
                     /evidence=not_experimental
     repeat_region   complement(3782. .3985)
                     /rpt_family="LTR40b"
                     /evidence=not_experimental
     exon            complement(3956. .4044)
                     /note="MZEF, score = 55.6%"
                     /evidence=not_experimental
     repeat_region   complement(3986. .5837)
                     /rpt_family="L1M1"
                     /evidence=not_experimental
     exon            complement(4289. .4459)
                     /note="GRAIL, score = 40.000%, comment = marginal"
                     /evidence=not_experimental
     exon            complement(5038. .5305)
                     /note="GRAIL, score = 63.000%, comment = good"
                     /evidence=not_experimental
     exon            5515. .5667
                     /note="MZEF, score = 83.3%"
                     /evidence=not_experimental
     repeat_region   complement(5905. .6209)
                     /rpt_family="AluYb8"
                     /evidence=not_experimental
     misc_feature    5950. .6070
                     /note="GC score = 5.40 (121bp)"
                     /note="Region: GC content"
                     /evidence=not_experimental
     repeat_region   complement(7126. .7473)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(7503. .8942)
                     /rpt_family="MIRA4"
                     /evidence=not_experimental
     repeat_region   complement(10008. .10319)
                     /rpt_family="L1M3b"
                     /evidence=not_experimental
     exon            10079. .10238
```

```
                     /note="GRAIL, score = 69.000%, comment = good"
                     /evidence=not_experimental
     exon            11517..11573
                     /note="MZEF, score = 52.7%"
                     /evidence=not_experimental
     exon            complement(13112..13170)
                     /note="GRAIL, score = 60.000%, comment = good shadow"
                     /evidence=not_experimental
     exon            14138..14187
                     /note="GRAIL, score = 96.000%, comment = excellent"
                     /evidence=not_experimental
     exon            15247..15348
                     /note="GRAIL, score = 87.000%, comment = excellent"
                     /evidence=not_experimental
     repeat_region   complement(15770..15922)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   complement(16238..16532)
                     /rpt_family="L1MA7"
                     /evidence=not_experimental
     repeat_region   complement(17159..17464)
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     exon            complement(17671..17802)
                     /note="GenScan, score = 2.21%, comment = Terminal_exon 132
                     bp frame: 0 phase: 0"
                     /evidence=not_experimental
     exon            complement(17671..17802)
                     /note="GRAIL, score = 61.000%, comment = good"
                     /evidence=not_experimental
     exon            complement(18026..18316)
                     /note="MZEF, score = 93.3%"
                     /evidence=not_experimental
     exon            complement(18026..18217)
                     /note="GenScan, score = 8.10%, comment = Initial_exon 192
                     bp frame: 1 phase: 0"
                     /evidence=not_experimental
     exon            complement(18026..18119)
                     /note="GRAIL, score = 50.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   18434..18495
                     /rpt_family="MIR"
                     /evidence=not_experimental
     exon            complement(18464..18486)
                     /note="GRAIL, score = 51.000%, comment = good"
                     /evidence=not_experimental
     exon            complement(19304..19390)
                     /note="MZEF, score = 69.2%"
                     /evidence=not_experimental
     repeat_region   complement(19657..19963)
                     /rpt_family="AluY"
                     /evidence=not_experimental
     exon            complement(19877..19972)
                     /note="MZEF, score = 61.8%"
                     /evidence=not_experimental
     repeat_region   20097..20294
                     /rpt_family="MIR"
                     /evidence=not_experimental
     promoter        complement(20809..20848)
                     /note="GenScan, score = -4.25%, comment = Length 40 bp"
                     /evidence=not_experimental
     exon            20853..20898
                     /note="GRAIL, score = 57.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(22543..22924)
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            22696..22752
                     /note="GRAIL, score = 72.000%, comment = good"
                     /evidence=not_experimental
     exon            22946..22942
                     /note="GRAIL, score = 88.000%, comment = excellent"
                     /evidence=not_experimental
     repeat_region   24822..25781[illegible]
                     /rpt_family="L1ME"
                     /evidence=not_experimental
     repeat_region   25797..2628[illegible]
                     /rpt_family="L1MC4"
                     /evidence=not_experimental
     repeat_region   26267..2686[illegible]
                     /rpt_family="L1M4"
                     /evidence=not_experimental
     exon            27475..276[illegible]
                     /note="GRAIL, score = 70.0[illegible]%, comment = good"
                     /evidence=not_experimental
     promoter        28837..288[illegible]
                     /note="GenScan, score = -4.4[illegible]%, comment = Length 40 bp"
                     /evidence=not_experimental
     exon            complement(28878..28978)
                     /note="GRAIL, score = 80.0[illegible]%, comment = excellent"
                     /evidence=not_experimental
     exon            28892..2891[illegible]
                     /note="Xpound exon prediction, score = [illegible]% (+)"
                     /evidence=not_experimental
     repeat_region   29680..2975[illegible]
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   30584..307[illegible]2
                     /rpt_family="MIR"
                     /evidence=not_experimental

  Query Match             9.3%;  Score 35.2;  DB [illegible];  Length [illegible];
  Best Local Similarity   52.0%;  Pred. No. [illegible].4;
  Matches   79;  Conservative    [illegible];  Mismatches   [illegible];  Indels    [illegible];  Gaps    0;

Qy        219 tctttctctggaccaaagtgaa[illegible] 278
Db       6611 TTTTTTTAGTTACAAGAGGGAA[illegible]AGTATAAT 6670
Qy        279 tgaaattgacatcaatgtggat[illegible]gacta 338
Db       6671 AATAATAGTAATAAAAGAAGCT[illegible] 6730
Qy        339 cactggccatggtaagtcacta[illegible] 37[illegible]
Db       6731 CTTTTACCTTCTTACGCCATTA[illegible] 67[illegible]2


RESULT   7
AC004949
LOCUS       AC004949    172366 bp    DNA             PRI       12-JAN-1999
DEFINITION  Homo sapiens PAC clone [illegible]18N.9 from [illegible], complete sequence.
ACCESSION   AC004949
VERSION     AC004949.1  GI:4150936
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172366)
  AUTHORS   Lamar,B., Wohldmann,P., Langston,Y., [illegible] and [illegible],K.
  TITLE     The sequence of Homo sapiens PAC clone [illegible]
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 172366)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-19[illegible]) Genome Sequencing [illegible] Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 172366)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-199[illegible]) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 12, 1999 this sequence version replaced gi:[illegible].
            SUBMITTED BY:  WUGSC
```

Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/CIB/CHP7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION.
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com), or from Pieter de Jong
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 197 of DJ1008N09;
actual end is at 172366 of DJ1008N09.

FEATURES	Location/Qualifiers
source	1. .172366
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7p21-p22"
	/clone_lib="RPCI-5"
	/clone="DJ1008N09"
repeat_region	1 4579
	/rpt_family="L1"
repeat_region	4692. .4739
	/rpt_family="AT_rich"
repeat_region	5331. .5446
	/rpt_family="L2"
repeat_region	5911 5937
	/rpt_family="(TAA)n"
repeat_region	6081. .6125
	/rpt_family="AT_rich"
repeat_region	6177. .6212
	/rpt_family="AT_rich"
repeat_region	6213. .6510
	/rpt_family="Alu"
repeat_region	6512. .6562
	/rpt_family="(CAGA)n"
repeat_region	6826. .6982
	/rpt_family="MIR"
repeat_region	7529. .7663
	/rpt_family="MIR"
repeat_region	8594. .8654
	/rpt_family="AT_rich"
repeat_region	9111. .9146
	/rpt_family="AT_rich"
repeat_region	9165 9482
	/rpt_family="MER2_type"
repeat_region	9484. .9943
	/rpt_family "MER1_type"
repeat_region	9944. .9965
	/rpt_family="AT_rich"
repeat_region	10021 10169
	/rpt_family="MER1_type"
repeat_region	11361. .11386
	/rpt_family="(CA)n"
repeat_region	11708. .11750
	/rpt_family="AT_rich"
repeat_region	11787. .11832
	/rpt_family="AT_rich"
repeat_region	12336 12362
	/rpt_family="AT_rich"
repeat_region	12394 12557
	/rpt_family="MIR"
repeat_region	13986. .14111
	/rpt_family="L2"
repeat_region	14125. .14871
	/rpt_family="L2"
repeat_region	14874 .14896
	/rpt_family="(TGAA)n"
repeat_region	15074. .15098
	/rpt_family="AT_rich"
repeat_region	16334 16373
	/rpt_family="AT_rich"
repeat_region	16942. .17238
	/rpt_family="Alu"
repeat_region	17807. .17874
	/rpt_family="L2"
repeat_region	17875. .18185
	/rpt_family="Alu"
repeat_region	18186. .18361
	/rpt_family="L2"
repeat_region	19127 19168
	/rpt_family="AT_rich"
repeat_region	19211 19558
	/rpt_family="L1"
repeat_region	20421 20774
	/rpt_family="MaLR"
repeat_region	20775. .22333
	/rpt_family="MaLR"
repeat_region	22334. .22685
	/rpt_family="MaLR"
repeat_region	24053. .24101
	/rpt_family="(TGAA)n"
repeat_region	24102. .24214
	/rpt_family="L2"
repeat_region	24510. .24542
	/rpt_family="AT_rich"
repeat_region	27094 27193
	/rpt_family="MIR"
repeat_region	27958. .28007
	/rpt_family="(CA)n"
repeat_region	28633. .28780
	/rpt_family="AT_rich"
repeat_region	29526. .29596
	/rpt_family="AT_rich"
repeat_region	30819. .30890
	/rpt_family="L2"
repeat_region	32106. .32133
	/rpt_family="AT_rich"
repeat_region	32178. .32331
	/rpt_family "MIR"
repeat_region	32388. .32483
	/rpt_family="L2"
repeat_region	32977. .33279
	/rpt_family="MaLR"
repeat_region	33308. .33691
	/rpt_family="MaLR"

```
     repeat_region   33682. .33983
                     /rpt_family="Alu"
     repeat_region   33984. .34027
                     /rpt_family="MaLR"
     repeat_region   35193. .35214
                     /rpt_family="AT_rich"
     repeat_region   35215. .35486
                     /rpt_family="Alu"
     repeat_region   35625. .35975
                     /rpt_family="MaLR"
     repeat_region   36198. .36224
                     /rpt_family="AT_rich"
     repeat_region   36243. .36300
                     /rpt_family="AT_rich"
     repeat_region   36444. .36495
                     /rpt_family="AT_rich"
     repeat_region   38103. .38282
                     /rpt_family="MaLR"
     repeat_region   38507. .38546
                     /rpt_family="AT_rich"
     repeat_region   38875. .39251
                     /rpt_family="L1"
     repeat_region   39290. .44246
                     /rpt_family="L1"
     repeat_region   44247. .44545
                     /rpt_family="Alu"
     repeat_region   44546. .46046
                     /rpt_family="L1"

  Query Match              9.3%;  Score 85;  DB 40;  Length 172366;
  Best Local Similarity   54.2%;  Pred. No. 7.4;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy     182 atcctcggaaaatcaagcctcacagcccaaatatcaaatctttctctggaccaaagtgaag 241
           |.| |.|.|.|  | |.|||| | |.|     ||| ||  || |  |||  ||||
Db   65099 ATGCATGAAGAAAGTATGTAACAGACACACAAAGAAATATTATTCAGCCTTAAAAAGAAG 65158

Qy     242 gatctattctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatg 301
           |.  .|.|  |  |:| |    | |||      | |  |||  | | |  ||     |:
Db   65159 GAAAIGTTGTGGTTTATGGCAACATGGATGAAIGTGGAGAACATTATATTAAGTGAAATA 65218

Qy     302 aagttgatacc 312
           |.| | | |
Db   65219 AGCCAGACACC 65229


RESULT   8
ATAC004697
LOCUS       ATAC004697 113950 bp    DNA             PLN       07-AUG-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence,
            complete sequence.
ACCESSION   AC004697
VERSION     AC004697.1  GI:3402671
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
            euphyllophytes, Spermatophyta, Magnoliophyta, eudicotyledons, core
            eudicots, Rosidae, eurosids II, Brassicales, Brassicaceae,
            Arabidopsis.
REFERENCE   1  (bases 1 to 113950)
  AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.I.,
            Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113950)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 113950)
  AUTHORS   Rounsley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Aug 7, 1998 this sequence version replaced gi:2244...
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr.
            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone T16B24 is from Arabidopsis thaliana chromosome II and is contained
            in the YAC clone CIC11A6.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://genomic.stanford.edu/Chris/GENSCAN.html) and MZEF/GeneMark
            (http://www.ebi.ac.uk/genemark/mzef.html). Examples of the
            complete sequence aligned to protein sequences in Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/atdb.html).
            Annotated genes are submitted to GenBank for the annotation for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes with no significant protein
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity that are predicted by more
            than two gene prediction programs over their length are
            annotated as 'hypothetical' proteins. Genes predicting tRNAs are
            predicted by tRNAscan-SE (Lowe & Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/) and are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated or genes not having predicted
            exons by GRAIL are annotated as misc_features.
FEATURES             Location/Qualifiers
     source          1. .113950
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="CIC11A6"
                     /clone="T16B24"
     misc_feature    complement(1. .777)
                     /note="overlaps with BAC clone T16B24 (AC004677)."
     misc_feature    24. .479
                     /note="exon predicted by Xpressions, Inc. Probability
                     excellent_score=xxx"
     misc_feature    complement(1. .312)
                     /note="exon predicted by Xpressions, Inc. Probability xxx=xxx"
     misc_feature    complement(<751. .978)
                     /note="exon predicted by Xpressions, Inc. Probability xxx"
     misc_feature    1012. .1105
                     /note="exon predicted by Xpressions, Inc. Probability
                     excellent_score=xxx"
     misc_feature    complement(1255. .1541)
                     /note="exon predicted by Xpressions, Inc. Probability xxx"
     misc_feature    1635. .1664
                     /note="exon predicted by Xpressions, Inc. Probability
                     marginal_score=xxx"
     misc_feature    complement(1761. .1975)
                     /note="exon predicted by Xpressions, Inc. Probability xxx=lowexon"
     misc_feature    2562. .2678
                     /note="exon predicted by Xpressions, Inc. Probability xxx=lowexon"
     misc_feature    complement(2..2975)
                     /note="exon predicted by Xpressions, Inc. Probability marginal"
     misc_feature    complement(3056. .3245)
                     /note="exon predicted by Xpressions, Inc. Probability xxx=lowexon"
     misc_feature    3513. .3645
                     /note="exon predicted by Xpressions, Inc. Probability
                     marginal_score=xxx"
     misc_feature    complement(3711. .3817)
```

```
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    4131..4159
                     /note="exon predicted by xgrail, quality good"
     misc_feature    complement(5408..5496)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   complement(5580..5617)
                     /rpt_family="(GAA)n"
     mRNA            <5692..>7914
                     /gene="T16B24.1"
     gene            <5692..>7914
                     /gene="T16B24.1"
                     /note="F12L6.1"
     CDS             5692..7914
                     /gene="T16B24.1"
                     /codon_start=1
                     /product="putative white protein"
                     /protein_id="AAC29975.1"
                     /db_xref="GI:3402672"
                     /translation="MARIVAANDDDSMELNTISSIHDSTLGQLLKNVSDVRKMAIGDE
                     IPVHESLNQDYNDGYMRTVPPVLSFDNLTYNVSVPPKLDFPNLFFPPRETEDPEIAQTA
                     PPKTKTLLNNISGETRPGEIMAVLGASGSGKSTLIDALANRIAKGSLKGTVKLNGETL
                     QSKMLKVISAYVMQDDLLFPMLIVEETLMFAAELRLPRSLPKSKKKLRVQALIDQLGI
                     RNAAKIIIGDEGHRGISGGERRRVSIGIDIIHDPILLFLDEPTSGLDSTSAFMVVKVL
                     KRIAQSGSIVIMSIHQPSHRVIGLLDPLIFLSRGHIVYSGSPASLPPFFIEFGSPIPE
                     NENRTEFALDLIRELEGSAGGIPGLIEFNKKWQEMKKQSNRQPPLTPPSSPYPNLILK
                     EAIAASISRGKLVSGGESVAHGAIINTIILAVPAFANPMWIEIKTLSKRSMLNSRRQ
                     PELFGIRIASVVITGFILATVFWRLDNSPKGVQERLGFFAFAMSTMFYTCADALPVFL
                     QERYIFMRETAYNAYRRSSYVISHAIVSFPSLIFLSVAFAATTYWAVGLDGGLTGLLF
                     YCLIILASFWSGSSFVTFISGVVPSVMLGYTIVVAILAYFLLFSGFFINRNRIPDYWI
                     WFHYMSLVKYPYEAVLQNEFSDATKCFVRGVQIFDNTPLGELPEVMKLKLLGTVSKSL
                     GVTISSITCLTIGSDILRQQQVVQLSKWNCLFITVAPGFFFKILFYFILLLGSKNKRP
                     "
     mRNA            complement(join(8041..8412,8513..8742,8839..8971,
                     9254..9394,9753..9912,9940..10080,10162..10644,
                     10805..10973,11097..11194,11454..11555,11702..12121,
                     12423..>13328))
                     /gene="T16B24.2"
     gene            complement(8041..>13328)
                     /gene="T16B24.2"
     CDS             complement(join(8217..8412,8513..8742,8838..8971,
                     9254..9394,9753..9912,9940..10080,10162..10644,
                     10805..10973,11097..11194,11454..11555,11702..12121,
                     12423..13328))
                     /gene="T16B24.2"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAC29976.1"
                     /db_xref="GI:3402673"
                     /translation="MQNRYGVMGSQTQKYSYQYSTGSESAPWTGHSVENQAVENGNYS
                     NSNYYHPQPTGPATGNVQEIPNTVSFTISSTSGTANVAQDYSGYTPYQTSSDPHNYSN
                     TGYSNYYSGYQQQPSQSYPQFVGAYQNTGAPQPLSSFQNPGSYAGIPSYSGTYYNPAD
                     YQIAGGYQSINYNNQTAGSYPSINYSNQTPASNQGNYIDYTSNPYQNYTPDAANTHSS
                     TIATTPPVHYQQNYQQWTEYSNQTEVPCAPGTEKLSTPTTSAYSQSFPVPGVTSEMPA
                     SNSQPAPSYVQPWRPETDSSHPPSQQPGAAVSTSNDTYWMHQAPSLQAHHPVPPQNNY
                     QSPLETKPIYFTPFQGHQPATVPQEMNSQSSFHQAPLGYPQPTQTAPLVDSQPVSKVQ
                     IPTNPRIASNLPSGFTKMDKDSTAASAAQAPAYVSVSMPKPKDHTTAMSDPGTFPKSL
                     RGFVEPAFARCKDDKEKESCEVALRKIVKKAKEDNTLYTPDWDTEPLSTVTTTNVTNS
                     ESSSAQLSSLQNKSPTRRPKSWWEPLVEGKPFVKPASTFSSAVKFGVWNHQNENNKKS
                     SESFQKVDAATGFKPTYSGQNSAKKSFQFPVKPQFFSGGAATAIDDEASSDSDKDLTP
                     YYSSAMALAGSAEEKKRRDSRSKRFEKIQGHSPGNDLTKPKNANVGNLHSPRATALRL
                     SKVFDFSGSRAVFDIDWDAIAVKGTCQEIEKPYLRLTSAPDPATVPPEDVLEKALIMV
                     QDSQKNYLFKCLQLKSIKQDIVQFIHNHLIAKVYETHAPLALEAGDLPEYNQCLSQL
                     KTLYAEGVPGCSLEFAAYSLLYITLHSNNNPELLSSMSPLSEEDKKDFAVPHALSVRA
                     AVTSGNYVMFFRLYKTAPNMNSCLMDLYVEKMPYKAVNFMSRSCRPTIPVSYIVQVLG
                     FTGAASGIDEKEIDGMEDCGWLKTHGANIITDSNGDMLLDIKAISISLFMPEPEDA
                     VAHGDRNLDVNDFFIRT"
     misc_feature    complement(13999..14162)
                     /note="exon predicted by xgrail, quality excellent"
     repeat_region   14221..14343
                     /rpt_family="(TAAAA)n"
     repeat_region   14264..14368
                     /rpt_family="POLY_A"
     misc_feature    14396..14513
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    complement(14481..14593)
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    complement(14990..15018)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   complement(15076..15137)
                     /rpt_family="(TA)n"
     misc_feature    15131..15163
                     /note="exon predicted by xgrail, quality good"
     misc_feature    complement(15812..15825)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    15850..15869
                     /note="exon predicted by xgrail, quality marginal"
     misc_feature    15927..16004
                     /note="exon predicted by xgrail, quality good"
     misc_feature    16334..16424
                     /note="exon predicted by xgrail, quality marginal"
     mRNA            join(<16515..16715,16919..17366,17497..17672,
                     18047..>18301)
                     /gene="T16B24.3"
     gene            <16515..>18301
                     /gene="T16B24.3"
     CDS             join(16515..16715,16919..17366,17497..17672,18047..18301)
                     /gene="T16B24.3"

  Query Match              9.1%;  Score 34.4;  DB 8:  Length 11345;
  Best Local Similarity   48.9%;  Pred. No. 11;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    1;  Gaps    3;

Qy     191 gaatcaagctgacagccccaaaatatcaatttttctctgaaatcaatcatcaatgtcaatcaaaa 250
           |||| | |  || | ||  ||| |   |  ||| |||          ||  |||| |
Db   82855 GAATGATGATGCATAACCAGAAAATTTATTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 82914

Qy     251 tctctgatgataacttggacagtccagatgaaattgacatcaatgtgaatgaaattatta 310
           || | | | ||   | ||     |||| |  | | ||| ||||  |     |
Db   82915 TTTAAGTTCAGAAAACTGGCAAAAACCATGGCAAGAATATCAATTACTATTTTTTTTTTTT 82974

Qy     311 ccccgatgaagcagattctttgcgtaactgccatgtaatgaaatttttaa 370
           || |   |  ||| |||| | || ||    | | |  |  | | ||  | |||
Db   82975 ACCGCAGCCATCAAGAGTCTTCTATCAACCACAGCATTTTTTTATGCTAAATTTTTAAA 83034

Qy     371 gccaaagt 378
           |||||| |
Db   83035 ACCAAACT 83042


RESULT   9
MMSILY/c
LOCUS       MMSILY       1740 bp    mRNA            ROD       26-APR-1993
DEFINITION  M.musculus mRNA for stromelysin 1.
ACCESSION   X66402 S46808
VERSION     X66402.1  GI:296167
KEYWORDS    stromelysin.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1740)
  AUTHORS   Hammani,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1992) K. Hammani [illegible] Lab. de
            Chimie Physiologique-UCL 7539, Ave. Hippocrate, 75, B-1200
            Bruxelles, BELGIUM
REFERENCE   2  (bases 1 to 1740)
  AUTHORS   Hammani,K., Henriet,P. and Peckhout,Y.
  TITLE     Cloning and sequencing of a cDNA encoding mouse stromelysin 1
  JOURNAL   Gene 120 (2), 321-327 (1992)
  MEDLINE   93013057
```

```
FEATURES             Location/Qualifiers
     source          1. .1740
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /tissue_type="calvarium"
                     /clone_lib="cDNA"
                     /clone="pSTM-1"
     CDS             25. .1458
                     /note="EC_number=3.4.24.17 (pro)stromelysin 1"
                     /codon_start=1
                     /product="stromelysin-1"
                     /protein_id="CAA47029.1"
                     /db_xref="GI:296168"
                     /db_xref="MGD:MGI:97010"
                     /db_xref="SWISS-PROT:P28862"
                     /translation="MKSLPVLLWLCVVVCSSYPLHDSARDDDAGMELLQKYLENYYGL
                     AKDVKQFIKKKDSSLIVVKIGEMQKFLGLEMTGKLDSNTMELMHKPRCGVPDVGGFST
                     FPGSPKWRKSHITYRIVNYTPDLPRSVIDSAIEKALKVWEEVIPLTFSRISEGEADIM
                     ISFAVGEHGDFVPFDGPGIVLAHAYAPGPGINGDAHFDDDERWTEDVTGINLFLVAAH
                     ELGHSLGLYHSAKAEALMYPLYKSSTDLSRFHLSQDDVDGIQSLYGIPIASFDVLVVP
                     TKSNSLEPETSPMCSSILFFLAVSTLPGEVLFFKDPHFWPKSLPIPEPEFYLISSFWP
                     SLPSNMDAAYEVTNRDTVFIFKGNQFWAIRGHEELAGYPKSIHTLGLPAIVKKIDAAI
                     SNKEKPKTYFFVEDKYWRFDEKTQSMEPGFPKIAEDFPGVDSPVDAVFEAFGFLYFF
                     SGSSQLEFDPNAKKVTHILKSNSWFNC"
     sig_peptide     25. .75
                     /product="stromelysin precursor"
     mat_peptide     76. .1455
                     /note="EC_number=3.4.24.17"
                     /product="stromelysin-1"
     polyA_signal    1717 .. 1722
BASE COUNT      466 a    371 c    425 g    478 t
ORIGIN


  Query Match             9.1%;  Score 34.4;  DB 12;  Length 1740;
  Best Local Similarity   48.9%;  Pred. No. 8.7;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      88 gaaatggacatcccctttgaagagggcgtgctgagtcccaatgctgcagacatgaggcct 147
           |||| || ||   |||||| | | | | | | | | | | | | | | | | | | |||| |
Db    1034 GGAAGAGATGGCCAAAATGAAGAGATCAAATAAAATTCAGGTTCAGGAGTCCTGAGAGAT 975

Qy     148 gaacctcctaattctctggatcttaatgacattcatcctcgagaatcaagctcacagcc 207
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     974 TTGCGCCAAAAGTGCCTGTCTTTAAAAAACAGGACTTCTCCCCGGAGGGTGCTGACTGCA 915

Qy     208 cgaaatatcgaatctttctctggaccaaagtgtaggatctattctctctgatgataacttg 267
           |||| | |||  | |||| |  |  |  | || |  | ||  ||||| |
Db     914 TCAAAGAACAAAGTAGAGCTGCACATTGGTGATGTCTCAGGTTCCAGAGAGTTAGACTTG 855

Qy     268 gacagtcc 275
           ||  ||
Db     854 GTGGGTAC 847


RESULT  10
MMTRANS1/c
LOCUS       MMTRANS1     1959 bp    mRNA            ROD       25-FEB-1997
DEFINITION  M.musculus mRNA for transin 1.
ACCESSION   X63162
VERSION     X63162.1  GI:54871
KEYWORDS    metalloproteinase; transin-1 gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1858)
  AUTHORS   Li.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-1991) Li, Universitaet Bern, Abteilung f.
            klinisch-experimentelle Forschung, Tiefenaustr. 120, SWITZERLAND
REFERENCE   2  (bases 1 to 1858)
  AUTHORS   Li,F., Strange,R., Brice,A.R., Edwards,D.R., Murphy,G.,
            Saurer,S., Niemann,B. and Andres,A.C.
  TITLE     Expression of stromelysin 1 in TIMP-1 transgenic mammary
            gland and in early mammary tumors
  JOURNAL   Int. J. Cancer 59 (4), 516-526 (1994)
  MEDLINE   95048890
COMMENT     Related sequence: Chr.mkal 1.. Mu..us ... ... ... (1-48).
FEATURES             Location/Qualifiers
     source          1. .1858
                     /organism="Mus musculus"
                     /strain="Swiss"
                     /isolate="MI-24"
                     /db_xref="taxon:10090"
                     /tissue_type="[illegible]"
                     /cell_line="whole mouse mammary gland"
                     /clone_lib="[illegible] Zap II"
                     /clone="MI-24"
     gene            77. .1516
                     /gene="transin-1"
     sig_peptide     77. .136
                     /gene="transin-1"
     CDS             77. .1516
                     /gene="transin-1"
                     /codon_start=1
                     /protein_id="CAA44456.1"
                     /db_xref="GI:54871"
                     /db_xref="MGD:MGI:97010"
                     /db_xref="SWISS-PROT:P28862"
                     /translation="MKSLPVLLWLCVVVCSSYPLHDSARDDDAGMELLQKYLENYY
                     GLAKDVKQFIKKKDSSLIVVKIGEMQKFLGLEMTGKLDSNTMELMHKPRCGVPDVGGF
                     STFPGSPKWRKSHITYRIVNYTPDLPRSVIDSAIEKALKVWEEVIPLTFSRISEGEAD
                     IMISFAVGEHGDFVPFDGPGIVLAHAYAPGPGINGDAHFDDDERWTEDVTGINLFLVA
                     AHELGHSLGLYHSAKAEALMYPLYKSSTDLSRFHLSQDDVDGIQSLYGIPIASFDVLV
                     VPTKSNSLEPETSPMCSSILFFLAVSTLPGEVLFFKDPHFWPKSLPIPEPEFYLISSF
                     WPSLPSNMDAAYEVTNRDTVFIFKGNQFWAIRGHEELAGYPKSIHTLGLPAIVKKIDA
                     AISNKEKPKTYFFVEDKYWRFDEKTQSMEPGFPKIAEDFPGVDSPVDAVFEAFGFLY
                     FFSGSSQLEFDPNAKKVTHILKSNSWFNC"
     mat_peptide     380. .1513
                     /gene="transin-1"
BASE COUNT      544 a    379 c    447 g    488 t
ORIGIN


  Query Match             9.1%;  Score 34.4;  DB 12;  Length 1858;
  Best Local Similarity   48.9%;  Pred. No. 8.8;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      88 gaaatggacatcccctttgaagagggcgtgctgagtcccaatgctgcagacatgaggcct 147
           |||| || ||   |||||| | | | | | | | | | | | | | | | | | | |||| |
Db    1092 GGAAGAGATGGCCAAAATGAAGAGATCAAATAAAATTCAGGTTCAGGAGTCCTGAGAGAT 1033

Qy     148 gaacctcctaattctctggatcttaatgacattcatcctcgagaatcaagctcacagcc 207
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1032 TTGCGCCAAAAGTGCCTGTCTTTAAAAAACAGGACTTCTCCCCGGAGGGTGCTGACTGCA 973

Qy     208 cgaaatatcgaatctttctctggaccaaagtgtaggatctattctctctgatgataacttg 267
           |||| | |||  | |||| |  |  |  | || |  | ||  ||||| |
Db     972 TCAAAGAACAAAGTAGAGCTGCACATTGGTGATGTCTCAGGTTCCAGAGAGTTAGACTTG 913

Qy     268 gacagtcc 275
           ||  ||
Db     912 GTGGGTAC 905


RESULT  11
AC006281/c
LOCUS       AC006281   130037 bp    DNA             HTG       22-SEP-1999
DEFINITION  Plasmodium falciparum chromosome 12 clone [illegible], *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AC006281
VERSION     AC006281.4  GI:5919271
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
```

```
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
REFERENCE   1  (bases 1 to 130037)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Fowley,T., Tamaki,T., Kurdi,O.B.,
            Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 130037)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Sep 23, 1999 this sequence version replaced gi:4559570.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    46221: contig of 46221 bp in length
            *   46222    46431: gap of unknown length
            *   46432    90606: contig of 44175 bp in length
            *   90607    90806: gap of unknown length
            *   90807    99786: contig of 8980 bp in length
            *   99787    99986: gap of unknown length
            *   99987   104256: contig of 4270 bp in length
            *  104257   104456: gap of unknown length
            *  104457   111594: contig of 7138 bp in length
            *  111595   111794: gap of unknown length
            *  111795   130037: contig of 18243 bp in length.
FEATURES             Location/Qualifiers
     source          1..130037
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="3D7"
BASE COUNT    54069 a  11763 c  14018 g  49186 t   1001 others
ORIGIN


  Query Match             9.1%;  Score 31.4;  DB 41,  Length 130037,
  Best Local Similarity   56.0%;  Pred. No. 11;
  Matches   65;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy     232 aaagtaaaaaatatttttgtgatgatgaa.ttgaacagtcagatgaaaattgacatca 292
           |  ||||  |||  || | |   ||| ||  |  |  ||       |||||| ||||  |
Db   97580 AIAAIGAIGGAGAIAAIAAIAAIGAAGATGAIGATGAAGAAAATGATGAACTTGAGCTTG 97521

Qy     293 atgtggatgaaattgatacccccgatgaaaa.attatttttgaagtacaatggccat 349
           | | |||||||| |||   | ||||||| |  |||  |  |||| | | ||   ||
Db   97520 AGGATGATGAAGATGAGAGGAGAGAIGATGAIAAIGATGAAAAATGATGAT 97469


RESULT  12
AC007424/3
LOCUS       AC007424   200729 bp    DNA             HTG       21-OCT-1999
DEFINITION  Homo sapiens chromosome 12q22 103.4-106.5 clone RPCI11-1016B19, ***
            SEQUENCING IN PROGRESS ***, 124 unordered pieces.
ACCESSION   AC007424
VERSION     AC007424.17  GI:6091645
KEYWORDS    HIG; HIGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200729)
  AUTHORS   Murray,P.M., Adams,C., Bailey,J., Barbaria,J., Blankenburg,K.,
            Bodeau,P., Burke,T., Bowles,S., Brooks,A., Bihay,C., Burrac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferrarato,D.,
            Forcum-Tansey,J., Frantz,F., Ganesh,P., Gerrell,T.H., Gerrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hornes,M.,
            Holloway,C., Hosak,H., Jackson,L.F., Jaffer,I., Jiang,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
            Martin,F., Martinez,C., McLeod,M.P., Mei,S., Meiner,M., Morris,S.,
            Nash,S., Nelson,A., Nguyen,B., Nguyen,N., Nguyen,S., Oswal,G.,
            Parish,B., Paxton,S., Payton,B., Perez,L., Phillips,L., Pillos,M.,
            Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,B.,
            Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Taber,P.,
            Taylor,T., Vasquez,L., Vilson,R., Vo,Z., Wakefield,M., Washington,S.,
            Weinstock,G., Weinstock,J.S., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 21, 1999 this sequence version replaced gi:5543544.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 124 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      851: contig of 851 bp in length
            *     852     1703: contig of 852 bp in length
            *    1704     2586: contig of 883 bp in length
            *    2587     3418: contig of 832 bp in length
            *    3419     4309: contig of 891 bp in length
            *    4310     5135: contig of 826 bp in length
            *    5136     5932: contig of 797 bp in length
            *    5933     7301: contig of 1369 bp in length
            *    7302     8095: contig of 794 bp in length
            *    8096     9174: contig of 1079 bp in length
            *    9175    10013: contig of 839 bp in length
            *   10014    11107: contig of 1094 bp in length
            *   11108    12405: contig of 1298 bp in length
            *   12406    13647: contig of 1242 bp in length
            *   13648    14464: contig of 817 bp in length
            *   14465    15605: contig of 1141 bp in length
            *   15606    16406: contig of 801 bp in length
            *   16407    17292: contig of 886 bp in length
            *   17293    18094: contig of 802 bp in length
            *   18095    18940: contig of 846 bp in length
            *   18941    19833: contig of 893 bp in length
            *   19834    20826: contig of 993 bp in length
            *   20827    21691: contig of 865 bp in length
            *   21692    22076: contig of 385 bp in length
            *   22077    23217: contig of 1141 bp in length
            *   23218    24018: contig of 801 bp in length
            *   24019    24811: contig of 793 bp in length
            *   24812    25598: contig of 787 bp in length
            *   25599    26400: contig of 802 bp in length
            *   26401    27242: contig of 842 bp in length
            *   27243    28096: contig of 854 bp in length
            *   28097    28736: contig of 640 bp in length
            *   28737    29983: contig of 1247 bp in length
            *   29984    30999: contig of 1016 bp in length
            *   31000    32183: contig of 1184 bp in length
            *   32184    32970: contig of 787 bp in length
            *   32971    33653: contig of 683 bp in length
            *   33654    34379: contig of 726 bp in length
            *   34380    35455: contig of 1076 bp in length
            *   35456    36302: contig of 847 bp in length
            *   36303    37132: contig of 830 bp in length
            *   37133    38106: contig of 974 bp in length
            *   38107    38971: contig of 865 bp in length
```

```
*     38972    39823: contig of 952 bp in length
*     39824    40869: contig of 1046 bp in length
*     40870    41693: contig of 824 bp in length
*     41694    42594: contig of 901 bp in length
*     42595    43820: contig of 1226 bp in length
*     43821    45078: contig of 1258 bp in length
*     45079    45838: contig of 760 bp in length
*     45839    46793: contig of 955 bp in length
*     46794    47654: contig of 861 bp in length
*     47655    48735: contig of 1081 bp in length
*     48736    50086: contig of 1351 bp in length
*     50087    51503: contig of 1417 bp in length
*     51504    52454: contig of 951 bp in length
*     52455    53449: contig of 995 bp in length
*     53450    54601: contig of 1152 bp in length
*     54602    55410: contig of 809 bp in length
*     55411    56207: contig of 797 bp in length
*     56208    57050: contig of 843 bp in length
*     57051    57847: contig of 797 bp in length
*     57848    59270: contig of 1423 bp in length
*     59271    60492: contig of 1222 bp in length
*     60493    61298: contig of 806 bp in length
*     61299    62462: contig of 1164 bp in length
*     62463    63947: contig of 1485 bp in length
*     63948    65335: contig of 1388 bp in length
*     65336    66177: contig of 842 bp in length
*     66178    67176: contig of 999 bp in length
*     67177    68357: contig of 1181 bp in length
*     68358    69614: contig of 1257 bp in length
*     69615    70805: contig of 1191 bp in length
*     70806    72270: contig of 1465 bp in length
*     72271    73897: contig of 1627 bp in length
*     73898    75162: contig of 1265 bp in length
*     75163    76252: contig of 1090 bp in length
*     76253    77447: contig of 1195 bp in length
*     77448    78865: contig of 1418 bp in length
*     78866    80408: contig of 1543 bp in length
*     80409    81570: contig of 1162 bp in length
*     81571    82395: contig of 825 bp in length
*     82396    83743: contig of 1348 bp in length
*     83744    85349: contig of 1606 bp in length
*     85350    87529: contig of 2180 bp in length
*     87530    88653: contig of 1124 bp in length
*     88654    90087: contig of 1434 bp in length
*     90088    91464: contig of 1377 bp in length
*     91465    92297: contig of 833 bp in length
*     92298    93738: contig of 1441 bp in length
*     93739    94864: contig of 1126 bp in length
*     94865    96039: contig of 1175 bp in length
*     96040    97330: contig of 1291 bp in length
*     97331    99078: contig of 1748 bp in length
*     98079   100608: contig of 1530 bp in length
*    100609   101986: contig of 1378 bp in length
*    101987   103025: contig of 1039 bp in length
*    103026   104146: contig of 1121 bp in length
*    104147   105711: contig of 1565 bp in length
*    105712   106425: contig of 714 bp in length
*    106426   107862: contig of 1437 bp in length
*    107863   109144: contig of 1282 bp in length
*    109145   110795: contig of 1651 bp in length
*    110796   112601: contig of 1806 bp in length
*    112602   113346: contig of 745 bp in length
*    113347   114731: contig of 1385 bp in length
*    114732   116235: contig of 1504 bp in length
*    116236   117917: contig of 1682 bp in length
*    117918   119838: contig of 1921 bp in length
*    119839   121619: contig of 1781 bp in length
*    121620   123168: contig of 1549 bp in length
*    123169   126522: contig of 3354 bp in length
*    126523   129779: contig of 3257 bp in length
*    129780   132435: contig of 2656 bp in length
*    132436   133603: contig of 1168 bp in length
*    133604   137809: contig of 4206 bp in length
*    137810   141844: contig of 4035 bp in length
*    141845   155440: contig of 13596 bp in length
*    155441   169144: contig of 13704 bp in length
*    169145   182389: contig of 13245 bp in length
*    182390   206978: contig of 24589 bp in length
*    206979   244458: contig of 37480 bp in length
*    244459   275747: contig of 31289 bp in length
*    275748   300729: contig of 24982 bp in length
FEATURES             Location/Qualifiers
     source          1..300729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q22-1[illegible]"
                     /clone="RPCI11-[illegible]B1[illegible]"
BASE COUNT    87399 a  59819 c  [illegible] g  [illegible] t  [illegible] others
ORIGIN


  Query Match             9.1%;  Score 34.4;  DB 42;  Length [illegible];
  Best Local Similarity   63.1%;  Pred. No. [illegible];
  Matches   53;  Conservative    [illegible];  Mismatches   [illegible];  Indels   [illegible];  Gaps    0;

Qy     239 aaggatctattctctgatga[illegible] 298

Db  293711 AATGCTATTTTCAGTACATCA[illegible] 293652

Qy     299 atgaacttgataccccgatga[illegible]22

Db  293651 TTGATCTTGATCACCTGGATGA[illegible]28


RESULT 13
LLU90222
LOCUS       LLU90222    10066 bp    DNA             BCT       [illegible]-MAY-1999
DEFINITION  Lactococcus lactis plasmid pIL2614 replication protein (repB),
            orfX, type IC restri[illegible]
            subunit (hsdM), type IC [illegible] contains
            short-chain alcohol dehydrogenases motif, and phage abortive
            infection protein (abiF) genes, complete cds.
ACCESSION   U90222
VERSION     U90222.1  GI:2865239
KEYWORDS    .
SOURCE      Lactococcus lactis.
  ORGANISM  Lactococcus lactis
            Bacteria; Firmicutes; [illegible]occaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 8600)
  AUTHORS   Schouler,C., Clier,F., [illegible] and Chopin,M.C.
  TITLE     A type IC restriction-modification system in Lactococcus lactis
  JOURNAL   J. Bacteriol. 180 (2), [illegible] (1998)
  MEDLINE   98101482
REFERENCE   2  (bases 8450 to 10066)
  AUTHORS   Portugal,C., Clier,F., Valyasevi,R., [illegible] and Chopin,M.-C.
  TITLE     Characterization of [illegible] phage
            abortive infection
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 10066)
  AUTHORS   Chopin,M.-C., Clier,F., [illegible] Portugal,C.,
            Valyasevi,R. and Schouler,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1997) Genetique Microbienne, INRA, [illegible]
            Jouy-en-Josas 78352, France
FEATURES             Location/Qualifiers
     source          1..10066
                     /organism="Lactococcus lactis"
                     /plasmid="pIL2614"
                     /strain="IL[illegible]"
                     /db_xref="taxon:1358"
     repeat_region   404..471
                     /note="thr[illegible] repeat"
                     /rpt_type=tandem
                     /rpt_unit=4[illegible]
     gene            593..1753
```

```
                     /gene="repB"
     CDS             592. .1753
                     /gene="repB"
                     /note="RepB"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAC15894.1"
                     /db_xref="GI:2865240"
                     /translation="MEIISKNFRNSNEPTVCPLKEIEKPKIVEHNDLITSVAKMDKTP
                     LKIFELAVSCIDTENLPKDNTIYLSKEELFTFFDVSDNGKHSRFKEAISKMQEQAFFE
                     VREPKNKGFKFRRILPIPTVEWTDYDDKVMIRFNQDIMPYLIELKQNFIQYALSDLID
                     LDSKYSTIIYKWLSMNYNQYEHYSAKSGRPAEQVESYPNPSITVKELRVITDTVKDYP
                     EFINFFTWIIKFFIFFINAHEFNVTYDKIFKGPSIDSIVFHIEKKPTADENSYKLED
                     QAVIEDKFIKEETFKFINFEVMKSKYTLILLENMLLSFYEMTDTALMAGLQAHVYPLY
                     DELPSLRCINGVKTHISYVAKFIEDYSFPNVAKYLKKAIEQYLFTVKPQDLS"
     CDS             1750. .2429
                     /gene="orfx"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC15895.1"
                     /db_xref="GI:2865241"
                     /translation="MSEPLKTIKELADELGISKKYLQNKISYEKKKQNLSIGNWKKIS
                     GKNTIVISDKFQTLIKSWINSIEDKQEDKADSLEDKQEDKADSLEDKQEDKADSLEDK
                     QEDKADSLEDKQEDKADSLEDKQEDKADSLEDKQEDKADSLEDKQEDKANSLEDKQED
                     KADSLEDKQFDKFKSYFKEPISVLEKFLDKEQISNSELKKIIQQTQNLLDQQQRLALQ
                     DKKLLEEYKSEINELKALKMAQEDMKDGSSIPGEAQEEIEPLKAQLKLSEEEPNKPKE
                     KEPVKTESKKWWQLWK"
     gene            2639. .5716
                     /gene="hsdR"
     CDS             2639. .5716
                     /gene="hsdR"
                     /note="HsdR"
                     /codon_start=1
                     /transl_table=11
                     /product="type IC restriction subunit"
                     /protein_id="AAC15896.1"
                     /db_xref="GI:2865242"
                     /translation="MSHSEQMIENQFIQILSEKENQWIYRPDLKSEEALQQNFRGHLN
                     RINIAVLFEQLLIDKEFKQVKVEFSFLIGIPFLASQWLRGENGVAQVLLEREDGEKVT
                     LEAFRNKDISGGTSSYEVVHLVVPDSSPVDPGDVSLLINGLPIIHIELQKESAKDGFM
                     QAYYQIQRYAFDGFFKGIYATTQIMVIPNKVDIRYFAPPSEDIAEAYAPMKKFLFNWR
                     TEDNQTVSDLFDFTRTVLRIPDAHELISQYTILVDDPKNPKFLMALRPYQIHAIPKIP
                     PFAAQHFDIFIWHATGSGKTLTSFVATKLLAQNAIGVDRIVMVVPRTILAQTQEEFI
                     KPASEYHTCQTTGNSVANTLIVGTKNQKQLARNLLSSKNNNTILVTTIQKLSAAMFSA
                     QQESEEKGSNQFEKLPQEHIMFIVDEAHRAVSDEEMKPIKKILPNSIWFGLTGTPIFE
                     ANKKQENGTFARTISQQYGPQLHSYTITKNAMDDGAVLGFQVEYHSLISEEDLEVIVIQ
                     LNKGKIPGDAIQQFFLLPAEVEKDEHIRTMLQKIFNPRSVVKKFKVKNGFPIMSAIL
                     TIHSIAQAKHIYRILKPMKKNSILLNGPQFDRPHPLIDKDFPPVAIIFSINPDPLEYN
                     EQDDELVPIMKEYAKQFDASPYQDEKLYNQNINKPLARKEKQYQSDGQWLDFVIVVDP
                     LLTGFDSPAILILYIDRENNVLKLLQAFSRTNRIYTGKDSGLIVSFPKPFTMKENVQN
                     TFRLFSNENQNFDQLIPREYEEVKKEFIECSTLYKQSEADLSDNPHDLKTMIAQVSAY
                     QKLGKSYKAFPSYIQYEEDFEAFSEVVEQLPQYPGKTENVKTKIKEMIEDEEHPEEDF
                     EKILQFIAFSSQLNATHKDVVDSFYINQLLFAIQLNEAGAVEKFEKEIQQKDPQTQKM
                     YHTLKDQLVNIISEEIDVAQLKEISIQNEIQKLLQKEAEEFGLSFDFLQSAMNEYQGDK
                     KAIPYITHLLDSMILSKEEFFPKAGEKYFPPKVLEERLPQNFEQIQKWKEEL"
     gene            5716. .7311
                     /gene="hsdM"
     CDS             5716. .7311
                     /gene="hsdM"
                     /note="HsdM"
                     /codon_start=1
                     /transl_table=11
                     /product="type IC modification subunit"
                     /protein_id="AAC15897.1"
                     /db_xref="GI:2865243"
                     /translation="MATGINLLWASADILPGKMDASEYKNYLLDLIFYKYLSPAQLR
                     EVYEQENGKTDTFPERSTQYAGFMEWYEEDKDDLIENIQPKQGYFIQPDQLFYSYRIK
                     ADNYEFNLIDLQAGFNELERQGEEFSGLFADIDLNSIKLGSNALLRNVIIIEVLRALD
                     EIDLFEHNGDVIGDAYEYLIGEFASSAGKKAGEFYTPQAVSKIMSEITSIGQETRAPF
                     HIYDFAMGSGSLMLNIRPYLNNIEQVHYHGQELNTTIFNLARMNLILHGIDKEPMNLN
                     NGDTLDADWPSDEPYDFDFVMTNPPYSAKWSAADQFLSDPRFEPFGKLAPKSKADFAF
                     LLHGFYHLKEGGTMGIVLPHGVLFRGAAEGTIRQALLEMGAIDAVIGLPANIFFGTSI
```

```
                     PITVILLKRNKPSRKSLFLGAFQDIEPRKDKNKLELKDIKVSGTFRAYHV
                     ASFDEIQENDFNLNIPRYVDTFEEEEPVDLVAVNTNLLKINEDLVQTNVLLSMIDNF
                     AESEENQALIESMRLLLRGGHDE"
     gene            7301. .8518
                     /gene="hsdS"
     CDS             7301. .8518
                     /gene="hsdS"
                     /note="HsdS"
                     /codon_start=1
                     /transl_table=11
                     /product="type IC specificity subunit"
                     /protein_id="AAC15898.1"
                     /db_xref="GI:2865244"
                     /translation="MNISPKSFQLRFEGFTIEWEERKFTFNNPFAVCFNKLEYSPKQV
                     LSVAQMKLNPSNRNELLYMPLYNVLRFDLTALELNPSFTTAFTFVLLQIGIVSH
                     VFYVYPPICKMCTFMIVYINNESVMFYLLVKATTKTLMTTLNTPTTVVPKLNLPSL
                     EEQQKIISFFKQLDAIALQKLDLFRLKKSLMQQLFV/RLRFAGFADDW
                     EDRKLGELASFSPSNGYIKNDLVEFGDRIIYDKLYKPTGLPPVVCPNNKFDKSII
                     SGGSEVIVPASCFSMELRAAVVGKSIILDDFINIFKYNGIPLNALTISNGSQ
                     QKEMSKRAQGKSVVLMPNSLKQVNILYSKLCFQLKPDKPFKLMNTVHQPKLDFL
                     KEQKKGFLQKMFV"
     terminator      8563. .8604
     CDS             8642. .9223
                     /note="orf1: contains short-chain alcohol dehydrogenases
                     motif"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC15899.1"
                     /db_xref="GI:2865245"
                     /translation="MNNVFKPIKYLSIIMLVVIISAVITMVVNMFMTLNSVSSNIS
                     KSVSQKSGLNLVVAYIFNNGFKVFMGMLFFSIIPIKFLWIQMLPIAILPGILFGIAF
                     FYSVAKAFILLFSLFRMLLEIFAFCLWMVALFFNPKMIIPPLVPPLINTPLFYEFK
                     LILISYVKYVIPLIVIAAFIEIYVADWISHILS"
     gene            9239. .9973
                     /gene="abiP"
     CDS             9239. .9973
                     /gene="abiP"
                     /note="AbiP"
                     /codon_start=1
                     /transl_table=11
                     /product="phage abortive infection protein"
                     /protein_id="AAC15900.1"
                     /db_xref="GI:2865246"
                     /translation="MILLPLEALINSLKLIPLINVSGLFFLGAFVYGAYSVMQ
                     KQLKVENEKYPKDFIENTFENLGLFQNVFFELFSSFITSLFKFFRKIGEDPYSIF
                     ISIDVNNIINKQDDIVLLNELVFKSNKNYSINYERALSPTLKSLIDSLKMEDKKEYS
                     GVLPGILSSKEMLVAFYNCMYFEKQFKEFKTLFVFFNTKSLSSLSFKNLDKGND
                     LPFFGKEDLLFSKTIDMQKLEELIKGN"
BASE COUNT     2656 a   1579 c   1969 g   2969 t
ORIGIN


  Query Match             9.0%;  Score 34.2;  DB 2;  Length 1[illegible];
  Best Local Similarity   54.3%;  Pred. No. 11;
  Matches   69;  Conservative    0;  Mismatches   58;  Indels    [illegible];  Gaps    0;

Qy     212 atatcaatctttctctggaccaaaatgaaaggatctattctctctgatggatgaaaaaaaaaa 271
           ||| | | ||||| | || |||    || | ||||      ||| | | | | | |
Db    7010 ATAAAATTGTTTCTATCCACAAAAAAACGAGAAGATATTGAAAGATATGCTCATCTGATGAAA 7069

Qy     272 gtccagatgaaaattgacatcaatgtggatgaacttgatacccccctatgaaacaaattttt 331
           ||   ||||| || |  |||      | | | ||| |  |  |  |  |  |  |  |
Db    7070 GTTTTGATGAGATCCAAGAAAATGACTTTAACTTAAATATCCCTCGTTTATCTATATTT 7129

Qy     332 ttgagta 338
           ||||| |
Db    7130 TTGAGGA 7136


RESULT  14
ATI4D2/c
LOCUS       ATI4D2      92611 bp    DNA             PLN       [illegible]-[illegible]-99
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T4D2.
```

```
ACCESSION   AL132958
VERSION     AL132958.1  GI:6424245
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 92611)
  AUTHORS   Nyakatura,G., Hartmann,B., Dauner,D., Sterr,W., Holland,R.,
            Weichselgartner,M., Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F.
            and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92611)
  AUTHORS   EU Arabidopsis sequencing,project
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1..92611
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
BASE COUNT    29024 a  16957 c  16763 g  29867 t
ORIGIN


  Query Match             9.0%;  Score 34.2;  DB 8;  Length 92611;
  Best Local Similarity   53.2%;  Pred. No. 12;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0,  Gaps    0;

Qy     219 atctttctctggaccaaagtgaaggatctattctctctgatgataacttggacagtccag 277
           ||  || |   | | ||| |||| | ||||  |||| |||||   |  ||  |  |
Db   53952 ATACTTATAAAGTCGAAAATGAATAAGTTATTAACTCTTTTGATATTGTTCACTTTTGAA 53893

Qy     278 atgaaattgacatcaatgtggatgaacttgataccccgatgaagcagattcttttgagt 337
           ||| |  | | |  |||  |||| || |    || ||| |  | |||  | || ||
Db   53892 ATGTAGATTATAGAAAATTGCGATGGTCTAAAAGGGTCCAAATGCAATATATTATGTTAAGC 53833

Qy     338 acactggccatggta 352
           | |   |  || | |
Db   53832 ATATATGTAATAGGA 53818


RESULT  15
AC012566/c
LOCUS       AC012566   172531 bp    DNA             HTG       03-DEC-1999
DEFINITION  Homo sapiens chromosome 4 clone RP11-38D12 map 4, WORKING DRAFT
            SEQUENCE, 2 ordered pieces.
ACCESSION   AC012566
VERSION     AC012566.2  GI:6614022
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172531)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-38D12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172531)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,P., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Dalglish,A.,
            Cooke,P., DeArellano,K., Dewar,K., Doctolero,M., Dokhanian,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Park,E., Gage,I.,
            Galagan,J., Gardyna,S., Grant,J., Hagos,B., Howland,A., Horton,L.,
            Howland,J.C., Johnson,E., Jones,C., Kamal,M., Kanagas,A., Klein,C.,
            Lehoczky,J., Lui,J., Maddox,K., Mahadevan,M., Mastrian,S.N.,
            McEwan,P., McCurk,A., McKernan,K., Mehalko,M.E., Morii,M.J.,
            Morrow,J., Naylor,J., Nelmer,C.H., Nguyen,T., Nguyen,T.P.,
            Peterson,K., Pollara,V., Riley,P., Roy,A., Santhanam,V., Savery,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Thangavelu,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,T., Wang,C., Wei,X.,
            Wyman,D., Ye,W.C., Zhang,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 3, 1999 this sequence version replaced gi:6424245.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997
            http://ftp.genome.wi.mit.edu/pub/software/RepeatMasker.htm.
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.genome.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            ---------- Project Information
                Center project name: RP11
                Center clone name: RP11_11
            ---------- Summary Statistics
                Sequencing vector: M13; M77815; 1.0% of reads
                Chemistry: Dye-terminator Big Dye; 1.4 fold coverage
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162851 bases at least 40
                Consensus quality: 169869 bases at least 30
                Consensus quality: 171998 bases at least 20
                Insert size: 215000; agarose-ip
                Insert size: 172531; end-of-contigs
                Quality coverage: 8.0 in 320 bases; agarose-ip
                Quality coverage: 7.0 in 320 bases; end-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given. However the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession numbers will be preserved.
            *       1    78380  contig of 78381 bp in length
            *                   gap of unknown length
            *   78381   172531  contig of 94151 bp in length.
FEATURES             Location/Qualifiers
     source          1..172531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-38D12"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    52774 a  30937 c  31076 g  57244 t
ORIGIN


  Query Match             9.0%;  Score 34.2;  DB 44;  Length 172531;
  Best Local Similarity   58.3%;  Pred. No. 12;
  Matches   60;  Conservative    0;  Mismatches   43;  Indels    0,  Gaps    0;

Qy     190 agaatcaagctcacagccccaaatatcaatcttctctgtgacaaatgtaaaaatgtt 249
           |  |||||||| |||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db 132375 AGTAACAAGCTCGAAGCCATAATAAAAAA.C.CCCAC.AAA.AAAA..C.AAA..A...A.. 132316

Qy     250 ctctctgatgataacttggacagttccaaatgaaaattgacatga 292
           |  ||  | | | |  |||  |  |  |  |  |  |  |  |
```

Db 132315 GGCTTTGCTGCTGAATTCTACCAAACATTTAAAGAAGAACTAA 132273

Search completed: February 9, 2000, 08:54:50
Job time: 16504 sec

OM of: US-09-174-937-8 to: N_Geneseq_36:* out_format : pfs

Date: Feb 8, 2000 11:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09174937/runat_03022000_085108_13316/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USEP=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-174-937-8
Query length: 126
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 103.050000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X39873	-	457.00	973.30	1.8e-46	1227	! Gastric cancer associated gene
N_Geneseq_36:X39877	-	457.00	977.51	1.0e-47	816	! Gastric cancer associated gene.
N_Geneseq_36:T89228	+	147.00	295.77	7.1e-09	1386	! Apoptosis-regulating protein N
N_Geneseq_36:X41211	-	147.00	295.77	7.1e-09	1386	! Human Nip2 encoding cDNA. Poly
N_Geneseq_36:V42332	-	76.50	141.01	2.9e-07	1322	! Faxilus involutus PhyA1 phytas
N_Geneseq_36:X12966	-	76.00	124.80	2.1e-07	6590	! Enterococcus faecalis genome
N_Geneseq_36:V74516	+	72.00	106.67	2.2e-10	10397	! Staphylococcus aureus contig
N_Geneseq_36:T29509	-	71.50	221.51	1.6e-24	3500	! Yeast calcineurin-interacting
N_Geneseq_36:X18167	-	71.50	125.91	2.0e-10	2284	! Human D2H protein coding seque
N_Geneseq_36:X24905	-	71.50	126.40	1.9e-10	2178	! Human elongation factor-2 kina
N_Geneseq_36:X24906	-	71.50	126.42	1.9e-10	2175	! Mouse elongation factor-2 kina
N_Geneseq_36:T59909	-	71.00	117.05	5.4e-16	4849	! Yeast transcription regulatory
N_Geneseq_36:V15893	+	71.00	134.32	7.0e-13	909	! Murine DMP1 transactivation dom
N_Geneseq_36:V15892	-	71.00	122.35	3.0e-09	2305	! Murine cyclin E transcription
N_Geneseq_36:V15893	-	71.00	122.35	3.0e-09	2303	! Murine cyclin E transcription
N_Geneseq_36:V30315	+	71.00	133.69	7.6e-13	967	! Bacillus thuringiensis MIS-5 in
N_Geneseq_36:V52198	-	71.00	107.85	2.6e-21	11931	! Streptococcus pneumoniae genom
N_Geneseq_36:T98717	+	71.00	136.49	5.3e-13	737	! DNA encoding a S. pneumoniae 30
N_Geneseq_36:T13635	+	70.00	80.64	6.8e+03	133894	! AcNPV genomic DNA clone 6, A
N_Geneseq_36:T13639	+	70.00	122.64	2.7e-07	2073	! AcNPV ORF 22, residues 13513-1
N_Geneseq_36:T89254	-	69.50	110.77	1.1e-18	5250	! cDNA for rat CASPR/p190, a lig
N_Geneseq_36:V59727	-	69.50	127.08	1.7e-11	1336	! Human secreted protein gene 53
N_Geneseq_36:V74475	-	69.50	98.07	7.3e-24	22243	! Staphylococcus aureus contig
N_Geneseq_36:X04444	-	69.50	127.98	1.6e-11	1236	! Yeast immunophilin FKBP48 cDNA
N_Geneseq_36:T62735	-	69.00	120.38	4.1e-08	2300	! Herbicide-resistant acetolacta
N_Geneseq_36:V19941	-	69.00	78.19	9.3e-03	137507	! KSHV long unique coding regi
N_Geneseq_36:V73909	-	69.00	92.27	1.5e-03	35100	! KSHV LUR DNA (nucleotides 35,
N_Geneseq_36:T03885	+	68.50	113.76	9.7e-19	3933	! Human mucosal lymphocyte-1 alp
N_Geneseq_36:T63575	-	68.50	114.14	9.3e-19	3791	! Chicken beta-actin promoter.
N_Geneseq_36:V86260	+	68.50	136.52	5.2e-13	433	! EST clone AQ5. New polynucleoti
N_Geneseq_36:Q67599	+	67.50	129.56	1.2e-10	688	! Retinoblastoma protein Ap15 5'
N_Geneseq_36:T89253	-	67.50	108.51	2.9e-18	5294	! cDNA for human CASPR/p190, a l
N_Geneseq_36:V43260	+	67.50	127.76	1.6e-11	819	! DNA encoding a human anion chan
N_Geneseq_36:V52295	-	67.50	106.53	2.4e-17	6415	! Streptococcus pneumoniae genom
N_Geneseq_36:V52194	-	67.50	100.19	5.5e-20	11864	! Streptococcus pneumoniae genc
N_Geneseq_36:V15896	+	67.00	126.92	1.8e-11	800	! Human cyclin D transcription fa
N_Geneseq_36:V15897	+	67.00	126.92	1.8e-11	800	! Human cyclin D transcription fa
N_Geneseq_36:V15898	+	67.00	126.29	1.9e-10	850	! Human cyclin D transcription fa
N_Geneseq_36:N80338	+	66.50	115.47	7.8e-16	2182	! Sequence encoding protein kina
N_Geneseq_36:Q39967	-	66.50	111.19	1.6e-08	3305	! PKC-alpha cDNA. New expression
N_Geneseq_36:Q79892	-	66.50	115.18	8.1e-09	2245	! DNA encoding a protein kinase
N_Geneseq_36:T15007	-	66.50	116.28	7.1e-09	2214	! Type III (alpha-type) rat
N_Geneseq_36:T63576	-	66.50	115.18	8.2e-09	2264	! Mouse H2Kb gene promoter.
N_Geneseq_36:T79062	-	66.50	112.72	1.1e-08	285	! Protein kinase A-binding
N_Geneseq_36:X22117	-	66.50	121.20	4.2e-08	217	! Human secreted protein ge

seq_name: N_Geneseq_36:X39873

seq_documentation_block:
ID X39873 standard; DNA; 1227 BP.
AC X39873;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, Ohata Y, Old LJ,
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132448/11.
PI New isolated cancer associated nucleic acids and polypeptides -
PI isolated using sera from cancer patients, used to monitor progress
PT for the diagnosis, monitoring, treatment of cancers
PS Claim 67; Page 579-580; 787pp; English.
CC The invention relates to a method for diagnosing a disease characterised
CC by expression of a human cancer associated antigen recognised by
CC a nucleic acid molecule (NAM). The method comprises, eg, contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of the expression
CC product complexed with an HLA molecule, and (b) detecting the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods may be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and, encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 1227 BP; 388 A; 270 C; 304 G; 264 T;

alignment_scores:
Quality: 457.00 Length:
Ratio: 5.022 Gaps:
Percent Similarity: 97.849 Percent Identity: 97.619

alignment_block:
US-09-174-937-8 x X39873

Align seq 1/1 to: X39873 from: 1 to: 1227

24 LeuLeuSerGluAspValGlyMetAspIleProPheGlyGlyValIle 67
882 CTACTGTCAGAGGATGTAGGAATCGATATCCCTTTCGAAGAGGTTATC 931

40 uSerProSerAlaAlaAspMetArgProGluProProAspSerLeuAspL 87
932 GAGTCCCAGTGCTGCAGACATGACGCTTGAACCTCCTGATAGTCTAGAT 981

87 gLuAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 73
982 TTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCTCCAAATATCAAC 1031

```
      74 LeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsn.LeuA 90
         ||||||||||||||||||||||||||||||||||||||||||||||||| ||||
    1032 CTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTTGG 1081

      90 spSerProAspGluIleAspIleAsnValAspGluLeuAspThrProAsp 106
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1082 ACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCGGAT 1131

     107 GluAlaAspSerPheGluTyrThrGly 115
         |||||||||||||||||||||    |||
    1132 GAAGCAGATTCTTTTGAGTACCCTGGC 1158

seq_name: N_Geneseq_36:X39876

seq_documentation_block:
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen, diagnosis, research, treatment, human,
KW   breast cancer, colon cancer, gastric cancer, renal cancer, lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999
PF   15-JUL-1998; 98WO-US014879.
PR   22-JUN-1998; 98US-089822.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132449/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 98; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    131 C;    209 G;    184 T;

alignment_scores:
            Quality:  457.00            Length:     93
              Ratio:  5.022               Gaps:      1
 Percent Similarity:  97.849   Percent Identity: 97.849

alignment_block:
 US-09-174-937-8 x X39876   ..

 Align seq 1/1  to: X39876  from: 1  to: 816

      24 LeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValLe 40
         |||||||||||||||||||||||||||||||| ||||||||||||||||
     471 CTACTGTCAGAGGATGTAGGAATGGACATCCCTTTGAAGAGGGCGTGCT 520

      40 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 57
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     521 GAGTCCCAGTCGTCGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATC 57

      57 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 7
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     571 TTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATATCAAT 620

      74 LeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsn.LeuA 9
         |||||||||||||||||||||||||||||||||||||||||||||||| ||||
     621 CTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTTGG 670

      90 spSerProAspGluIleAspIleAsnValAspGluLeuAspThrProAsp 10
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     671 ACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCGGAT 720

     107 GluAlaAspSerPheGluTyrThrGly 115
         |||||||||||||||||||||    |||
     721 GAAGCAGATTCTTTTGAGTACCCTGGC 747

seq_name: N_Geneseq_36:T38228

seq_documentation_block:
ID   T38228 standard; cDNA; 1386 BP.
AC   T38228;
DT   18-DEC-1996  (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA
KW   Apoptosis-regulating protein; Nip2; adenovirus E1B 19K protein;
KW   Bcl-2; oncogene; cell death; cancer; tumour; immune disorder;
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214  .1158
FT                   /*tag=  a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PP   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G,
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998
PT   Nucleic acids encoding apoptosis regulating proteins - useful in
PT   diagnosing and treating immune disorders, malignancies, etc.
PS   Claim 1; Page 39-41; 60pp; English.
CC   6 cDNA clones (T38227-32) respectively code for novel apoptosis
CC   regulating proteins Nip1, Nip2, Nip3, BipIA, Bip5 and Bip6
CC   (W00997-W01002).  Nip clones were isolated from human cDNA
CC   libraries using a yeast two-hybrid strategy with adenovirus E1B
CC   19K protein (W01010) as bait; Bip clones were isolated using
CC   Bcl-2 oncoprotein (W01018) as bait.  E1B 19K and Bcl-2 both protect
CC   cells against apoptosis.  Nip1, Nip2, Nip3 bind to 19K, BipIA and
CC   Bip5 bind to 19K and Bcl-2, and Bip6 binds to Bcl-2 also.  The
CC   acids can be used in diagnostic assays, to produce recombinant
CC   apoptosis-regulating proteins, or to regulate cell death by
CC   increasing or suppressing expression of an apoptosis-blocking gene
CC   such as bcl-2.
SQ   Sequence   1386 BP;    428 A;    249 C;    336 G;    373 T;

alignment_scores:
            Quality:  147.00            Length:     89
              Ratio:  2.333               Gaps:      5
 Percent Similarity:  70.787   Percent Identity: 43.82

alignment_block:
 US-09-174-937-8 x T38228   ..

 Align seq 1/1  to: T38228  from: 1  to: 1386

      27 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerPro Se 4
         ||||||      ...  ..:|||   ||||||||||||   ||| ||
     250 GAAGATTTTCCGATACCTTTACCACAAGATCATAGTATTGAAGCAGATAT 2
```

```
      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         |||          ||||||      |||:::|||||||:::::|||
     300 ACTAGCTATAACTGGACCAGAGGACCAGCCTGGCTCACTAGAAGTTAATG 349

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::     |||:::   ||||||   |:|:|:::|||:::|||:::
     350 GAAATAAAGTGAGAAAG.  AAACTAATGGCTCCAGACATTAGCCTGACA 396

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||    |||:::|||||||:::|||||||:|:||    |||||||:::
     397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 443

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         :|||||||||:::      |||   |||||:|||||||::|||   .
     444 TGGGGAGATTGACTTA......GATGGCTTAGACACACCGTCAGAGAATA 487

     108 laAspSerPheGluTyr 113
         :::::::|||||:::
     488 GTAATGAGTTTGAGTGG 504
```

seq_name: N_Geneseq_36:X04120

seq_documentation_block:
ID X04120 standard; cDNA; 1386 BP.
AC X04120;
DT 13-APR-1999 (first entry)
DE Human Nip2 encoding cDNA.
KW Human; Nip1; Nip2; Nip3; Bip1A; Bip5; Bip13; adenovirus; cell death;
KW viral infection; Bcl-2; protooncogene; mutational analysis; apoptosis;
KW E1B 19K protein; cell survival regulation; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 214. .1158
FT /*tag= a
PN US5858678-A.
PD 12-JAN-1999.
PF 21-MAR-1995; 408095.
PR 02-AUG-1994; US-284139.
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 99-152099/13.
DR P-PSDB; W94324.
PT Polypeptides that bind to anti-apoptotic proteins - useful for
PT protecting against cell death induced by viral infection and to
PT modulate response to physical and chemical stimuli
PS Claim 2; Column 51-54; 41pp; English.
CC The present invention describes: (1) a method for regulating cell death,
CC comprising exposing an isolated cell to a polypeptide selected from
CC Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13; (2) a method for neutralising
CC the activity of the adenovirus E1B 19 kD protein, the Bcl-2 protein or
CC the BHRF-1 protein, comprising exposing an isolated cell to a
CC polypeptide as in (1), and (3) a method for detecting molecules that
CC bind to at least one polypeptide as in (1), comprising lysing cells,
CC exposing the lysate to the polypeptide and detecting any molecule-
CC polypeptide aggregates. The methods are useful for providing proteins
CC able to bind to other proteins known to regulate cell survival e.g. it
CC is known that E1B 19K protein provides a survival function similar to
CC the cellular protooncogene bcl-2 gene product which is able to block
CC apoptosis in haematopoietic B and T cells. The present sequence encodes
CC human Nip2, given in the present invention.
SQ Sequence 1386 BP; 428 A; 249 C; 336 G; 373 T;

alignment_scores:
Quality: 147.00 Length: 89
Ratio: 2.333 Gaps: 5
Percent Similarity: 70.787 Percent Identity: 43.820

alignment_block:
US-09-174-937-8 x X04120 ..

Align seq 1/1 to X04120 from 1 to 1386

```
      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 41
         ||:|||    :::  ::::|          :::::::::::::   :::
     250 GAAGATTTTGCGATACCTTTACCAGAAGATCAATGTATTGAAGCAGATAT 299

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         |||          ||||||      |||:::|||||||:::::|||
     300 ACTAGCTATAACTGGACCAGAGGA.AAGCCTGGCTCACTAGAAGTTAATG 349

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::     |||:::   ||||||   |:|:|:::|||:::|||:::
     350 GAAATAAAGTGAGAAAG...AAACTAATGGCTCCAGACATTAGCCTGACA 396

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||    |||:::|||||||:::|||||||:|:||    |||||||:::
     397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 443

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
         :|||||||||:::      |||   |||||:|||||||::|||   .
     444 TGGGGAGATTGACTTA......GATGGCTTAGACACACCGTCAGAGAATA 487

     108 laAspSerPheGluTyr 113
         :::::::|||||:::
     488 GTAATGAGTTTGAGTGG 504
```

seq_name: N_Geneseq_36:V42332

seq_documentation_block:
ID V42332 standard; cDNA; 1522 BP.
AC V42332;
DT 09-NOV-1998 (first entry)
DE Paxillus involtus PhyA1 phytase cDNA.
KW Phytase; PhyA1; basidiomycetes; feed additive; food additive; ss.
OS Paxillus involtus strain CBS 208.91.
FH Key Location/Qualifiers
FT CDS 58. .1386
FT /*tag= a
FT sig_peptide 58. .114
FT /*tag= b
FT mat_peptide 115. .1383
FT /*tag= c
PN WO9828409-A1.
PD 02-JUL-1998.
PF 15-DEC-1997; DK0568.
PR 01-DEC-1997; DK-001388.
PR 23-DEC-1996; DK-001480.
PR 20-DEC-1996; DK-001481.
PR 18-MAR-1997; DK-000301.
PR 07-MAY-1997; DK-000529.
PA (NOVO ; NOVO-NORDISK AS.
PI Bech L, Breinholt J, Fuglsang CC, Lassen SF, Tomari A;
DR WPI; 98-377641/32.
DR P-PSDB; W62859.
PT Phytase(s) from fungi of phylum Basidiomycota - useful as feed and
PT food additives, e.g. to reduce phosphate content of manure and to
PT improve digestibility
PS Claim 25; Page 121-122; 197pp; English.
CC This cDNA sequence codes for the PhyA phytase (see W62859) of the
CC basidiomycete Paxillus involtus CBS 208.91. A plasmid containing
CC the full-length cDNA sequence in Stabble vector pYES 2.0 is
CC deposited in Escherichia coli DSM 11912. The cDNA was originally
CC isolated by expression cloning in yeast. The invention provides
CC basidiomycete phytases (see W62857-62), cloned DNA sequences (see
CC V42330-35), processes for preparing the phytases, and their use
CC especially as food or feed additives to improve digestibility.
CC Compared with known phytases, the claimed enzymes show better heat
CC stability or faster release of phosphate from substrates, and may
CC also show advantageous position specificity and specific activity.
CC They can be produced in commercially useful quantities.
SQ Sequence 1522 BP; 304 A; 484 C; 392 G; 342 T;

alignment_scores:

```
           Quality:   76.50              Length:     79
             Ratio:  1.663                Gaps:      4
 Percent Similarity:  59.228   Percent Identity: 32.911

alignment_block:
 US-09-174-937-8 x V42332   ..

 Align seg 1/1  ... V42332  from: 1  to: 1522

       5 CysSerArgAlaSerPheSerProSerValArgLysProProLeuIleLe 21
         |||||| |.|||||||:::::   |||   |||:::|||      :::::
     827 TGTTCGAGGGCATCCCTGGCTCTTTCGAGGCGTTCGCCT......ATGGT 870

      21 uArgArgLeuLeuSerGluAspValGlyMetAspIle          P 34
         :     |||   |||:::   |||   :::   :::          |
     871 GGGGACCTTGACAAGTTCTACGGTACCGATTA-GTCAGGAACTCGGACC 920

      34 roPheGluGluGlyValLeuSerProSerAlaAlaAspMetArgProGlu 50
         ||'|::::   ::::::   |||   ||'::::::          |||
     921 CGTTCAAGGCGTCGGCTACGTCAACGAGCTCA'CGCCGGCTCACCAACT 970

      51 ProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysIe 67
         :| |:|::::::               |||    ||||||||
     971 CGGCCGTCCGGC               ACAACACCCAGACGAACCGCACA 1005

      67 uThrAlaProAsn IleAsnLeuSerLeuAsp'n 78
         ::::::|||:::    ||||||   ::::::| |
    1006 CTCGACGCCTCGCCCGTAACCTTCCCGTTGAA.AA 1040

seq_name: N_Geneseq_36:X12969

seq_documentation_block:
ID   X12969 standard; DNA; 6590 BP.
AC   X12969;
DT   19-MAR-1999  (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO 32
KW   Enterococcus faecalis contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; U08985.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT     used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 392 395; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence   6590 BP;    1980 A;    1414 C;    1065 G;    2134 T;


alignment_scores:
           Quality:   76.00              Length:     77
             Ratio:  1.854                Gaps:      4
 Percent Similarity:  53.247   Percent Identity: 38.766

alignment_block:
 US-09-174-937-8 x X12969   ..

 Align seg 1/1  to: X12969  from: 1  to: 6590

       6 SerArgAlaSerPheSer    .              ProSerValArgLysPr 17
         |||:::|||||||||                     |||:::::   |'|
    5410 TCTAAAGCAAGCTTTTGCACACCTTGGCGGTTACCAACAATCATTAAGGA 5461

      17 oProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMetAspIleP 34
         |||      ::     |||   ::::::   ::|||          :
    5460 TCCGGGAGTAATATATCGCGTCATCGCGGTTTCTGTCATCGTAGCAATTA 5511

      34 roPheGluGluGlyValLeuSerProSerAlaAlaAspMetArgProGl 51
                    :::|||||                  | ' || |
    5510 CAAATTTATTAAGGACTTTGTCT.................AGACCAGAG 5541

      51 ...ProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLy 66
           |||||||||                 |||   |||       | ::
    5542 CTACCGCCTAAC ..              ACATCGCCTTCAATAATCCG 5601

      66 sLeuThrAlaProAsnIleAsnLeuSerLeu 76
         :  |||:::|||||||:::|||::: |':::
    5574 GACCACTTCACCAAATGTTAATTTTCAAIA 5604

seq_name: N_Geneseq_36:V74516

seq_documentation_block:
ID   V74516 standard; DNA; 16397 BP.
AC   V74516;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #205.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection, surgical wound infection, scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    421. .480
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    2221. .2280
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    4021. .4080
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    5821   5880
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    7621. .7680
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    9421. .9480
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
```

```
FT                              the sequence listing in the specification. They
FT                              are included to maintain the nucleotide numbering
FT                              given in the specification for this DNA sequence"
FT   misc_feature    11221. .11280
FT                   /*tag= g
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    13021. .13080
FT                   /*tag= h
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    14821. .14880
FT                   /*tag= i
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009961.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 922-931, 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   16397 BP;    4397 A;    3179 C;    2525 G;    5744 T;


alignment_scores:
            Quality:   72.00              Length:     76
              Ratio:  1.636                Gaps:      1
 Percent Similarity:  57.895   Percent Identity: 23.684

alignment_block:
 US-09-174-937-8 x V74516   ..

 Align seq 1/1  to: V74516  from: 1  to: 16397

       6 SerArgAlaSerPheSerProSerValArgLysProProLeuIleLeuAr 22
         |||:::          ..:..:|||||||:::|||:::   |||   :::::::||
   10747 TCAAATTTGATGTATGAACCGICAIIACGACGAACACCTGACTTAGTACG 10796

      22 gArgLeu. .. .. .. .. .. .. .. .. ..LeuSerGluAspValGlyM 31
         |   :::                          | |:::         |||:
   10797 TACGATTACAGCTTTGACAACGTCACCIIIIIIAACAACGCCACCTGGTG 10846

      31 etAspIleProPheGluGluGlyValLeuSerProSerAlaAlaAspMet 47
```

```
         ::   :::          :::   :::     :  :                 :::
   10847 TTGCATTTTTAACAGTACATACCATAACATCGIIGATCTTTGIIIIIIA 1 ...

      48 ArgProGluProProAsnSerLeuAspLeuAsnAspIleHisIleArgAr 64
         | | |:::|||||| | |:::     :::::::   :::::::     ::
   10897 CGACCAGATCCACCTAATACTTTGATTGTAAGAACTICACGAATATCATA 1...

      64 gIleLysLeuThrAlaProAsnIleAsn 73
         :::::::   |||          ::::::::::
   10947 GTTGTCTGCTACTTTCAAGCGTGTTIII 1197

seq_name: N_Geneseq_36:T18500

seq_documentation_block:
ID   T18500 standard; DNA; 3500 BP.
AC   T18500;
DT   27-OCT-1996  (first entry)
DE   Yeast calcineurin-interacting polypeptide CNI DNA.
KW   Calcineurin-interacting polypeptide; CNI; immunosuppressive;
KW   immunosuppressant; gene therapy; ds.
OS   Yeast.
FH   Key             Location/Qualifiers
FT   cds             376. .1320
FT                   /*tag= a
PN   WO9612806-A1.
PD   02-MAY-1996.
PF   23-OCT-1995; U13580.
PR   24-OCT-1994; US-328322.
PA   (STPD ) UNIV LELAND STANFORD JUNIOR.
PI   Cyert MS, Huang L;
DR   WPI; 96-230607/23.
DR   P-PSDB; T18500.
PT   Calcineurin interacting polypeptide - enhances immunosuppressive
PT   effects of calcineurin-targetted immunosupressant by potentiating
PT   interaction between an immunophilin and CN
PS   Disclosure; Page 42-46; 90pp; English.
CC   This DNA encodes CNI, a calcineurin-interacting polypeptide. This
CC   nucleic acid may be used in expression vectors for gene therapy to
CC   potentiate the immunosuppressive effects of immunosupresant drugs.
CC   This DNA may be expressed recombinantly to prepare a CNI protein
CC   and may also be used to form a fusion protein which consists of a CN
CC   subunit and this CNI polypeptide.
SQ   Sequence   3500 BP;    1135 A;    754 C;    [illegible] T;


alignment_scores:
            Quality:   71.50              Length:     [illegible]
              Ratio:  1.430                Gaps:      4
 Percent Similarity:  49.505   Percent Identity: 28.7[illegible]

alignment_block:
 US-09-174-937-8 x T18500   ..

 Align seq 1/1  to: T18500  from: 1  to: 3500

      36 GluGluGlyValLeuSerProSerAlaAlaAspMet.....Ar 4[illegible]
         | ::::   ::::|||||:::   ||| :::   ::::::
     814 GAAGATTATCTACTTGAAAAAAGCIIIIGTTTCAITATATATTTTCATCAAA 863

      49 gProGluProProAsnSerLeuAspLeuAsnAspIleHisIleArgAr 95
         |:|||||||:::::
     864 CGCTGAACCCCCTAAAGAT .  ..... ... 8[illegible]

      65 leLysLeuThrAlaProAsnIleAsnLeuSerLeuAspIleSer[illegible]Gly 81
                                |:|                 |
     883 ..................AATAGCAATCTAAGTTTGGATICCTATCAATTTG 916

      82 SerIleLeuSerAspAspAsnLe  .AspSerProAspGluIleAspIl 1[illegible]
         :::   |.||||:::||.|||||     ::::::    :::    ::
     916 AACTCATTGTCATCCGATAATTT[illegible]AATAACAAIICCATTATTCAAGTTTTTT 975

      97 eAsnValAspGlu.....................[illegible]Asp[illegible] 1[illegible]
         ::::   ' ::::                                 :
```

```
     966 TTCCCATGACGACAGTAAACTGTTGGCTATTCAAAAGACACCATTACCAT 1015

     104 hrProAspGluAlaAspSerPheGluTyrThrGlyHisGlyLysSerLeu 120
         ::   ::::::      |||          :::   ||||||::||||||
    1016 CATCTAGTCGAAGAGGATCGGTACCGGCAAATTTTCACGGTAACTCTTTG 1065

     121 Ser 121
         |||
    1066 TCA 1068

seq_name: N_Geneseq_36:X18167

seq_documentation_block:
ID   X18167 standard; DNA; 2284 BP.
AC   X18167;
DT   04-MAY-1999  (first entry)
DE   Human D2H protein coding sequence.
KW   Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW   D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW   intestinal peptide-associated transporter; hypertension; diabetes;
KW   osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW   therapeutic agent delivery; therapy; probe; ss.
OS   Homo sapiens.
PN   WO9851325-A2.
PD   19-NOV-1998.
PF   15-MAY-1998; 98WO-US009588.
PR   15-MAY-1997; 97US-046595.
PA   (CYTO-) CYTOGEN CORP.
PA   (ELAN-) ELAN CORP PLC.
PI   Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ,
PI   Omahony DJ, Patterson CA, Singleton J;
DR   WPI; 99-009568/01.
DR   P-PSDB; W74091.
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and
PT   antibodies, used to deliver therapeutic or diagnostic agents to, or
PT   through, the gastrointestinal tract, e.g. insulin or leuprolide
PS   Disclosure; Fig 4; 294pp; English.
XX
CC   This sequence encodes the human D2H protein. The invention relates to
CC   purified proteins (I) that bind specifically to at least one of the
CC   gastro-intestinal (GI) tract receptors human intestinal
CC   peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC   sucrose-isomaltase complex (hSI). (I) provide active transport of
CC   therapeutic agents through human and animal GI tissue (into the blood)
CC   for in vivo delivery, particularly for treatment or prevention
CC   of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC   migraine, or angina pectoris. Specifically they are used to deliver
CC   insulin or leuprolide, but many other suitable therapeutic agents are
CC   disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC   antigens. (I) may also provide targeting to the GI tract. Other uses of
CC   (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
SQ   Sequence  2284 BP;  690 A;  485 C;  514 G;  595 T;


alignment_scores.
            Quality:   71.50            Length:    145
              Ratio:  1.067              Gaps:      9
 Percent Similarity:  46.207   Percent Identity: 27.586

alignment_block:
 US-09-174-937-8 x X18167/rev   ..

 Align seq 1/1 to reverse of X18167  fr.  1  to  2284

       5 CysSerArgAlaSerPheSerProSerValArgLysProProLeuIleLe 21
         |||||||:::   |||   ||||||      ||:||   |||||||:::
    1993 TGTTCAAAGATGAGTCCCTCTCCCTTGTCCAGAAAAATGCCACTTGTA.. 1946

      21 uArgArgLeuLeuSerGluAspVal yMetAspIleProProGlu   5F
                                                        :
    1945 . ... ..... . ..................................  1905

      38 lyValLeuSerProSerAlaAlaAspMet..................... 67
         ....:|||   |||   ...|||::::::
    1944 CAACTTTACTGCCTTTGTCGGCAGAAATGGTACTTAACCTTATTCTTAT  1895

      48 .       ArgProGlu.     .......ProProAsnSerLeuAsp.  57
                 |||||||||                  ||||  :::
    1894 TTAGCGGGAAGGCCCGAAATCATATTATGTAGATTTAACAGTGTTGATTC 1845

      57 euAsnAspThrHisProArg...ArgIleLysLeuThrAlaProAsnIle 72
         ||:..:.....:||||||   ||:::          :||:   :::
    1844 TCCAAAATTCAGAACCACGATAAACATTCTGTCGATGCCATCCAGCTCTC 1795

      73 AsnLeuSerLeuAspGlnSerGlyGlySerIleLeuSerAspAspAsnLe 89
                 |||     ::::::::::::     ||:::::::::::
    1794         TTGTGTACACAACATAGTGGCTGTCATTCCTCAAATCCAA 1754

      89 uAspSerPro    .          .          AspGluIleAspIleAsnV 98
           ||||||                          |||::|||.:::|||
    1753 AACCAGCCCCTGTTGAGGAGTAGCTCATTGGCATGAAGTAGACTTAAATG 1704

      99 alAspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGly 115
         ::|||      :::   |||      ||:::::||||||        :
    1703 TTGATATAACTTCAAAGCCGATCTGGGCTGAGTCTTTT... ...  35A 1664

     116 HisGlyLysSer            LeuSerTrp 122
         |||       | |            :::: ||
    1662 CATCAACATTCACAGTGTGGTAATCTGAATTGG 1630

seq_name: N_Geneseq_36:X24905

seq_documentation_block:
ID   X24905 standard; cDNA; 2178 BP.
AC   X24905;
DT   21-JUN-1999  (first entry)
DE   Human elongation factor-2 k... ...EF-2 kinase; cDNA.
KW   Elongation factor-2 kinase; eEF-2 kinase; human; protein kinase;
KW   inhibitor; breast cancer; therapy; ds.
OS   Homo sapiens.
PN   WO9909199-A2.
PD   25-FEB-1999.
PF   20-AUG-1998; 98WO-US017272.
PR   20-AUG-1997; 97US-914999
PA   (UYNE-) UNIV NEW JERSEY.
PI   Hait WN, Pavur KS, Ryazanov AG;
DR   WPI; 99-181050/15.
DR   P-PSDB; W98106.
PT   New isolated protein kinase, eEF-2 - used to develop assays to
PT   controlling the amount, or activity of protein kinases, e.g. for
PT   treating cancers or other hyperproliferative pathologies
PS   Claim 18; Page 104-105; 195pp; English.
XX
CC   This cDNA clone encodes human elongation factor-2 kinase (eEF-2
CC   kinase, see W98106), a member of a new superfamily of eukaryotic
CC   protein kinases that phosphorylate within an alpha-helical domain
CC   of a target protein, as opposed to beta-turns as seen in all other
CC   protein kinases. eEF-2 kinase cDNA was obtained from human glioma
CC   cell line T98G using primers based on the mouse eEF-2 kinase coding
CC   region (see X24906). eEF-2 kinase is a ubiquitous enzyme which is
CC   involved in the regulation of protein synthesis and the cell cycle.
CC   It has no homology to any other calmodulin kinase, and is
CC   therefore an ideal target in the search for a specific protein
CC   kinase inhibitor. Since preliminary evidence suggests that eEF-2
CC   kinase is upregulated in human cancers, including breast cancer,
CC   identification of specific inhibitors of eEF-2 kinase may lead to
CC   the development of novel anti-cancer drugs. Assays have been
CC   developed utilising eEF-2 kinase and a phosphorylation target
CC   (see W98109) to facilitate high throughput screening for small molecules
CC   that can specifically inhibit eEF-2 kinase. Methods of assessing
CC   eEF-2 kinase levels for diagnostic purposes, and therapeutic
CC   formulations to inhibit eEF-2 kinase activity are also disclosed.
```

```
CC   Sequences complementary to eEF-2 kinase may have therapeutic
CC   efficacy as antisense drugs or be used in gene therapy.  A
CC   ribozyme that cleaves eEF-2 kinase mRNA is also claimed.
SQ   Sequence   2178 BP;    541 A;    677 C;    637 G;    323 T;


alignment_scores:
            Quality:   71.50              Length:    148
              Ratio:  1.192                Gaps:      8
 Percent Similarity:  40.541   Percent Identity: 25.676

alignment_block:
 US-09-174-937-8 x X24905   ..

 Align seq 1/1  to: X24905  from: 1  to: 2178

       1 MetLeuLysSerCysSerArgAlaSerPheSerProSerValArg      15
         ||||:|:::::   ::::::          |  :|   ||||||
    1045 ATCTTGAGAGGAACAGAGGAAAAAATGTGGGAGTCCCCGAGTAAGGACCCT 1094

      16 .........LysProProLeuIleLeuArgArgLeuLeuSerGluAspV 29
                 ::::::::::::::::     |::   ||||||||||:::
    1095 CTCTGGGAGCCGGCCACCCCTGCTC......CTCCCCTTTCAGAGAACT 1138

      29 alGly............MetAspIleProPheGluGluGlyValLeuSer 41
         |  |             | ::::     ::::::       |||
    1139 CTGGAGACGAGAACATGAGCGACGTGACCTTCGACTCTCTCCCTTCTTCC 1188

      42 ProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspLeuAs 58
         :|||| ::: ||                  |||:::|||     ||||::
    1189 CCATCTTCGGCC..............ACACCACACAGCCAGAAGCTAGA 1223

      58 nAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuS 75
         :   ::
    1224 CCACCTCCAT..................................... 1233

      75 erLeuAspGlnSerGluGlySerIleLeuSerAsp............... 86
                     ::::::    |
    1234 .................TGGCCAGTGTTCAGTCACCTCGATAACATGGCA 1266

      87 .......     AspAsnLeuAspSerProAspGluIleAspIleAsnVa 99
                  ||:::'''|| ...  ||   ...  ...
    1267 TCCAGAGACCATGATCATCTAGACAACCACCTCGAGTCTGAGAAIAGTGG 1316

      99 lAsp............................GlnLeuAspThrProAspG 107
         | |                            :  ||||||   |||:::
    1317 GGACAGCGGATACCCCAGTGAGAAGCGGGGGTGAGCTGGATGACCCTGAGC 1366

     107 luAlaAspSerPheGluTyrThrGlyHisGlyLysSerLeuSer 121
            :::              |||' '   |||   |||
    1367 CCCGAGAA          ....CATGGCTACTCATACAGT 1392

seq_name. N_Geneseq_36.X24905

seq_documentation_block:
ID   X24905 standard; cDNA; 2178 BP.
AC   X24905;
DT   21-JUN-1999  (first entry)
DE   Mouse elongation factor-2 kinase (eEF-2 kinase) cDNA.
KW   Elongation factor-2 kinase; eEF-2 kinase; mouse; protein kinase;
KW   inhibitor; breast cancer; therapy; ds.
OS   Mus musculus.
PN   WO9909199-A2.
PD   25-FEB-1999.
PF   20-AUG-1998; 98172?2.
PR   20 AUG 1997; US-914999.
PA   (UYNE-) UNIV NEW JERSEY.
PI   Hait WN, Pavur KS, Ryazanov AG;
DR   WPI; 99-181050/15.
DR   P-PSDB; W98107.
PT   New isolated protein kinase, eEF-2 - used to develop agents for
PT   controlling the amount, or activity of protein kinases, e.g. for
PT   treating cancers or other hyperproliferative pathologies

PS   Claim 19; Page 109-110; 195pp; English.
CC   This cDNA clone encodes mouse eEF-2 kinase (eEF-2
CC   kinase; see W98107), a member of a novel eukaryotic protein
CC   protein kinases that phosphorylate a hydroxyl amino acid in their
CC   of a target protein, as opposed to beta-turns as seen in all other
CC   protein kinases.  eEF-2 kinase cDNA was obtained from mouse spleen
CC   cDNA using primers based on rabbit eEF-2 kinase sequence. It
CC   was used to identify a human eEF-2 kinase clone (see X24906).
CC   eEF-2 kinase is a ubiquitous enzyme which is involved in the
CC   regulation of protein synthesis and the cell cycle. It has no
CC   homology to any other mammalian protein kinase, and is therefore an
CC   ideal target in the search for a specific protein kinase inhibitor.
CC   Since preliminary evidence suggests that eEF-2 kinase is
CC   upregulated in human cancers, including breast cancer,
CC   identification of specific inhibitors of eEF-2 kinase may lead to
CC   the development of novel anticancer drugs. Assays have been
CC   developed utilising eEF-2 kinase and a phosphopeptide substrate
CC   (see W98109) to facilitate high-throughput screening for compounds
CC   that can specifically inhibit eEF-2 kinase. Methods for assessing
CC   eEF-2 kinase levels for diagnostic purposes and therapeutic
CC   formulations to inhibit eEF-2 kinase activity are also disclosed.
CC   Sequences complementary to eEF-2 kinase mRNA, can be used in gene
CC   efficacy as antisense drugs or be used in gene therapy.  A
CC   ribozyme that cleaves eEF-2 kinase mRNA is also claimed.
SQ   Sequence   2178 BP;    548 A;    [illegible]


alignment_scores:
            Quality:   71.50              Length:    [illegible]
              Ratio:  1.067                Gaps:      4
 Percent Similarity:  49.268   Percent Identity: 36.471

alignment_block:
 US-09-174-937-8 x X24906   ..

 Align seq 1/1  to: X24906  from: 1  to: 2178

       1 MetLeuLysSerCysSerArgAlaSerPheSerProSerValArg..... 17
         :::   :::::::   ::::::                  :::
    1042 ATCTTGAGGGGGACAGAGCAGAAATTTTTTATTCCTTAAAAATACATT 1 1

      18 ..........LysProProLeuIleLeuArgArgLeuLeuSerGluAsn 34
                 :::  :   :::                     :::
    1092 CTCTAGCAGCCGGCCCCCTTGCTCTTTT.....CTCTTATATAAAT 11+

      29 alGly............MetAspIleProPheGluGlyValLeuSer 41
         |  |             | :::     ::::::
    1136 CCGGGGATGAGAACATGAGTGACGTATCTTTATCTTCTTTTTTTTT 11+

      42 ProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspLeu... ... 57
                 :::  :   :::     :::
    1186 CCGTCTTCAGTTACACCACACAGCCAGAAAATTGACTATTTCATTTTTT 12+

      56 ..........AspLeuAsnAspThrHisPr ArgArgIleLysLeuThrA 69
                 '''  : ::::::
    1236 AGTGTTTGGTGACCTCGATAACATTTTTTATA..................... 12+

      69 laProAsnIleAsnLeuSerLeuAsp TleSerGluGlyArgIleLeuSer 87
                    :::     :::   ::::::
    1270 ..................GACCATTTTTTATTCTTAATTATTTTACTTT 13+

      88 AspAspAsnLeuAspSer.....Ile AspGlnIleAspIleAsnVal Asp 1
         ::::::::::   '''  :     :::     :::   ::
    1303 GAGAATAGTGGGGACAGTGGGTATTAAGTTATA.....AATTTAAT 134+

     100 pGluLeuAspThrProAspGluAlaAspSerProGlnTyrLeuSerGlu 117
         ::::    :   :  :::     :::
    1344 TGACCTGGATGATCCTGAGCCCCTATAA.....TAATT 1377

     117 lyLysSer 119
         '

    1376 GCCACTCC 1383
```

seq_name: N_Geneseq_36:I59909

seq_documentation_block:
ID I59909 standard; DNA; 4849 BP.
AC I59909;
DT 04-JUN-1997 (first entry)
DE Yeast transcription regulatory factor SRB9 DNA.
KW Transcription regulatory factor; suppressor of RNA polymerase B;
KW SRB9; RNA polymerase II holoenzyme; SWI/SNF; ss.
OS Saccharomyces cerevisiae.
FH Key Location/qualifiers
FT cds 148. .4410
FT /*tag= a
PN WO9708301-A1.
PD 06-MAR-1997.
PF 28-AUG-1996; U14192.
PR 31-AUG-1995; US-521872.
PR 11-OCT-1995; US-540804.
PR 26-JAN-1996; US-590399.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PI Chao DM, Kolyeke AJ, Thompson CM, Young PA;
DR WPI; 97-179258/16.
DR P-PSDB; W13826.
PT Purified RNA polymerase II holoenzyme - comprises RNA polymerase II
PT and one or more regulatory proteins, pref. suppressor of RNA
PT polymerase B proteins or SWI/SNF proteins
PS Claim 11; Fig 11a-c; 154pp; English.
CC Novel DNA sequences (I59904-11) code for yeast SRB (suppressor of
CC RNA polymerase B) proteins SRB4, SRB5, SRB6, SRB7, SRB8, SRB9,
CC SRB10 and SRB11 (W13821-28), respectively. SRBs are transcription
CC regulatory factors that act as positive and negative regulators
CC of RNA polymerase II activity. They are components of the RNA
CC polymerase II holoenzyme. The sequences were identified using
CC methods designed to identify transcription factors involved in RNA
CC polymerase II C terminal domain (CTD) function. The SRB9 allele
CC suppresses conditional phenotypes associated with a CTD truncation
CC mutation. SRB nucleic acids and RNA polymerase II holoenzymes can
CC be used to treat diseases resulting from alteration or deletion of
CC the SRB gene, pref. by gene transfer technology. Probes based on
CC the genes can be used to detect SRB nucleic acids.
SQ Sequence 4849 BP; 1611 A; 97 C; 923 G; 1345 T;

alignment_scores:
Quality: 71.00 Length: 150
Ratio: 1.000 Gaps: 6
Percent Similarity: 54.615 Percent Identity: 26.923

alignment_block:
US-09-174-937-8 x I59909 ..

Align seq 1/1 to: I59909 from: 1 to: 4849

10 PheSerProSerValArgLysProProLeu..... ..IleIle 21
1792 ITTTTTCCCTCATCTGAAAAACCGAATATCCG'ACGATGAGCGGAACTAC 1841

21 ArgArgLeuLeuSerGluAspValGlyMetAspIleProPheGluGlu 38
1842 AAAAAGATTAAATGGAAAGAGGAAATATTTGGATATTCCGATAGATGAAA 1891

38 lyValLeuSerProSerAlaAlaAspMetArg..oGluProProAsnSer 54
1892 T'ACTTGCCAACGAGTCCATTATATATGGATC'AGGTGCGCCA 1935

55 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaPro.. 70
1936 CTCCCTGTGGAAACACCCCGCCATAGACGCAAAACTGTGTTCGCTCCACT 1985

71 .AsnIleAsnLeuSerLeuAspGlnSer . . .GluG 81
1986 GAATTTTAACCCCATAATAGAAAACAATGTTGATAACAAATACAAATCTG 2035

81 lySerIleLeuSerAspAspAsnLeuAspSerProAspGluIle..... 95
2036 GACCCAAATTTTCCTTCAGTCCGTTCTAAAAGCAGGAAGCATTAAACTTT 2085

96 AspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerProGl 112
2086 GATATTTCTATGGCGATCTTTCTACTTCTCAACAGCAACGCATTAAAT 2135

112 uTyrThrGlyHisGly .....LysSerLeuSer 121
2136 AGAGAACGGTAGCAGCGATGAGGATTCAAATTCATTCAAC 2175

seq_name: N_Geneseq_36:V15893

seq_documentation_block:
ID V15893 standard; cDNA; 909 BP.
AC V15893;
DT 26-MAY-1998 (first entry)
DE Murine DMP1 transactivation domain C-terminal fragment encoding cDNA.
KW cyclin D transcription factor, binding affinity, D type cyclin;
KW cell cycle inhibitor, tumour detection, cancer, murine DMP1;
KW transactivation domain; ds.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 1. .909
FT /*tag= a
FT /product= "DMP1 transactivation domain"
FT /note= "fragment of V15882, does not contain start
FT and stop codons"
PN WO9743415-A1
PD 20-NOV-1997.
PF 16-MAY-1997; U08480.
PR 15-MAY-1997; US-017815.
PR 16-MAY-1996; US-017815.
PR 16-MAY-1996; US-648837.
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PI Hirai H, Inoue K, Sherr CJ;
DR WPI; 98-008884/01.
DR P-PSDB; W47096.
PT Cyclin D transcription factor and related DNA - can be used to
PT develop products for treatment of e.g. cancer
PS Claim 19; Pages 91-93; 120pp; English.
CC This cDNA encodes a transactivation domain of murine cyclin D
CC transcription factor DMP1. DMP1 is an amino acid polymer which has
CC binding affinity for a D-type cyclin, in vitro and for a specific
CC DNA nucleotide sequence and is a transcription factor involved in the
CC activation of genes that prevent cell proliferation. The DMP1 nucleic
CC acid is operatively linked to an expression control sequence in an
CC expression vector. The expression vector has a transcription control
CC sequence comprising a nonamer sequence (CCCGTATGT) operably associated
CC with a recombinant gene or a cassette insertion site for a recombinant
CC gene. The vector is homologously recombined in a chromosome of a
CC transgenic animal. The presence of activity of DMP1 can be determined by
CC detecting binding of DMP1 and a probe by contacting a biological sample
CC from a mammal with the probe under conditions that allow binding of the
CC probe to DMP1, where the probe contains the core sequence GTA, and where
CC the presence or activity of DMP1 is suspected in the sample. DMP1 can
CC function as a cell cycle inhibitor when expressed in a tumour cell.
CC Modulating the expression of DMP1 can be used to treat tumours and other
CC cancers. DMP1 can also be used for controlling expression of heterologous
CC proteins. Antisense sequences and ribozymes can be used to inhibit
CC expression of the transcription factor. Detecting the level and activity
CC of DMP1 in cells is useful for detection of cancer cells or
CC dysproliferative cells
SQ Sequence 909 BP; 273 A; 244 C; 156 G; 236 T;

alignment_scores:
Quality: 71.00 Length: 106
Ratio: 1.109 Gaps: 5
Percent Similarity: 60.377 Percent Identity: 25.472

alignment_block:

```
US-09-174-937-8 x V15893  ..

Align seq 1/1  to: V15893  from: 1  to: 899

     17 ProProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMetAspIl 33
        |||   :..|||  :..   :..  :.|||  :....     :::
    331 CCAAGAGTCATCATTCAGACGGTAGCTACAGAGGACATCACTTCTTCATT 380

     33 eProPheGluGluGlyValLeuSerProSer.....AlaAlaAspMetA 48
        :     |||||/   :.....:.    ...      :::::|||:::
    381 ATCCCAAGAGGAACTGACAGTTGATAGTGATCTTCATTCATCTGATTTC. 429

     48 rgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArg 64
        |||||||||||||:::::|||:::     |:|||   |||   :::
    430 ..CCTGAGCCTCCAGATGCACTAGAAGCA...GACACTTTCCCAGACGAA 474

     65 IleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGl 81
        |||                 :::|||:::.  :| :.:::::::::::
    475 ATTCCTCGGCCTAAGATGACTATACAACCATCATTTAATAATGCTCATGT 524

     81 ySerIleLeuSerAspAspAsn.........................L 89
        |||   :::|||||||:: ||                          :
    525 ATCTAAATTCAGCGACCAAAATAGCACAGAACTGATGAACAGTGTCATGG 574

     89 euAspSerProAspGluIle...AspIleAsnValAspGluLeuAspThr 104
        ::   :..   ..:|||||||   |||   :::...   :::   :::
    575 TCAGAACAGAGGAAGAAATTGCCGACACTGACCTTAAGCAGGAAGAACCG 624

     105 ProAspGluAlaAspSer 110
         |||::::::      |||
    625 CCGTCTGACTTAGCCAGT 642

seq_name: N_Geneseq_36:V15882

seq_documentation_block:
ID   V15882 standard; cDNA; 2903 BP.
AC   V15882;
DT   26-MAY-1998  (first entry)
DE   Murine cyclin D transcription factor DMP1 encoding cDNA
KW   cyclin D transcription factor; binding affinity; D-type cyclin;
KW   cell cycle inhibitor; tumour; detection; cancer; murine; DMP1; ds.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             249. .2533
FT                   /*tag=  a
FT                   /product= "DMP1"
PN   WO9743415-A1.
PD   20-NOV-1997.
PF   16-MAY-1997; 97WO-US008480.
PR   15-MAY-1997; 97US-017815.
PR   15-MAY-1996; 96US-017815.
PR   16-MAY-1996; 96US-648837.
PA   (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PI   Hirai H, Inoue K, Sherr CJ;
DR   WPI; 98-008884/01.
DR   P-PSDB; W47093.
PT   Cyclin D transcription factor and related DNA - can be used to
PT   develop products for treatment of, e.g. cancer
PS   Claim 21; Pages 80-81; 120pp; English.
CC   This cDNA encodes a murine cyclin D transcription factor DMP1, an amino
CC   acid polymer which has binding affinity for a D-type cyclin, in vitro,
CC   and for a specific DNA nucleotide sequence and is a transcription factor
CC   involved in the activation of genes that prevent cell proliferation. The
CC   DMP1 nucleic acid is operatively linked to an expression control sequence
CC   in an expression vector. The expression vector has a transcription
CC   control sequence comprising a nonamer sequence (CCCGTATGT) operably
CC   associated with a recombinant gene or a cassette insertion site for a
CC   recombinant gene. The vector is homologously recombined in a chromosome
CC   of a transgenic animal. The presence of activity of DMP1 can be
CC   determined by detecting binding of DMP1 and a probe by contacting a
CC   biological sample from a mammal with an oligonucleotide probe under
CC   conditions that allow binding of the probe to DMP1, where the nucleotide
CC   probe contains the core sequence GTA, and where the presence or activity
CC   of DMP1 is suspected in the sample. DMP1 can [illegible] be a cell cycle
CC   inhibitor when expressed in a tumour cell. Modulating the expression of
CC   DMP1 can be used to treat tumours and other [illegible]. DMP1 can also be
CC   used for controlling expression of heterologous proteins. Antisense
CC   sequences and ribozymes can be used to inhibit expression of the
CC   transcription factor. Detecting the level and activity of DMP1 in cells
CC   is useful for detection of cancer cells or dysplastic cells.
SQ   Sequence  2903 BP;  880 A;  662 C;  [illegible]

alignment_scores:
            Quality:  71.00            Length:    [illegible]
              Ratio:  1.109              Gaps:    [illegible]
 Percent Similarity:  60.377   Percent Identity:  [illegible]

alignment_block:
 US-09-174-937-8 x V15882  ..

 Align seq 1/1  to: V15882  from: 1  to: 2903

      17 ProProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMetAspIl ..
         ||:   :::| |||||:  :::  ::: :::  :::  |||||
    1952 CCAAGAGTCATCATTCAGACGGTAGCTACAGAGGACATCACTTCTTCATT 2.. 1

      33 eProPheGluGluGlyValLeuSerProSer.....AlaAlaAspMetA 48
         :     |||||/|   :::::::::  :::   :::::  :::
    2002 ATCCCAAGAGGAACTGACAGTTGATAGTGATCTTCATTCATCTGATTTC. 2.5.

      48 rgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArg 64
         |||||||||||||:::::|:|:::     :     :  :::
    2051 ..CCTGAGCCTCCAGATGCACTAGAAGCA...GACACTTTCCCAGACGAA 2.9.

      65 IleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGlu.. 81
         |||                 :::|:::.  :|:::::::::::::::
    2096 ATTCCTCGGCCTAAGATGACTATACAACCATCATTTAATAATGCTCATGT 2145

      81 ySerIleLeuSerAspAspAsn.........................L 94
         |||   :::|||||||:::|||                          :
    2146 ATCTAAATTCAGCGACCAAAATAGCACAGAACTGATGAACAGTGTCATGG 2195

      89 euAspSerProAspGluIle...AspIleAsnValAspGluLeuAspThr 104
         ::   :..   ..:|||||||   |||   :::...   :::   :::
    2196 TCAGAACAGAGGAAGAAATTGCCGACACTGACCTTAAGCAGGAAGAACCG 2245

     105 ProAspGluAlaAspSer 110
         |||::::::      |||
    2246 CCGTCTGACTTAGCCAGT 2263

seq_name: N_Geneseq_36:V15883

seq_documentation_block:
ID   V15883 standard; mRNA; 2903 BP.
AC   V15883;
DT   26-MAY-1998  (first entry)
DE   Murine cyclin D transcription factor DMP1 encoding mRNA.
KW   cyclin D transcription factor; binding affinity; D-type cyclin;
KW   cell cycle inhibitor; tumour; detection; cancer; murine; DMP1; ds.
OS   Mus musculus.
PN   WO9743415-A1.
PD   20-NOV-1997.
PF   16-MAY-1997; 97WO-US008480.
PR   15-MAY-1997; 97US-017815.
PR   16-MAY-1996; 96US-017815.
PR   16-MAY-1996; 96US-648837.
PA   (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PI   Hirai H, Inoue K, Sherr CJ;
DR   WPI; 98-008884/01.
DR   P-PSDB; W47093.
PT   Cyclin D transcription factor and related DNA - can be used to
PT   develop products for treatment of, e.g. cancer
PS   Claim 21; Pages 82-83; 120pp; English.
CC   This mRNA encodes a murine cyclin D transcription factor DMP1, an amino
CC   acid polymer which has binding affinity for a D-type cyclin, in vitro,
```

```
CC   and for a specific DNA nucleotide sequence and is a transcription factor
CC   involved in the activation of genes that prevent cell proliferation. The
CC   DMP1 nucleic acid is operatively linked to an expression control sequence
CC   in an expression vector. The expression vector has a transcription
CC   control sequence comprising a nonamer sequence (CCCGTATGT) operably
CC   associated with a recombinant gene or a cassette insertion site for a
CC   recombinant gene. The vector is homologously recombined in a chromosome
CC   of a transgenic animal. The presence of activity of DMP1 can be
CC   determined by detecting binding of DMP1 and a probe by contacting a
CC   biological sample from a mammal with an oligonucleotide probe under
CC   conditions that allow binding of the probe to DMP1, where the nucleotide
CC   probe contains the core sequence GTA, and where the presence or activity
CC   of DMP1 is suspected in the sample. DMP1 can function as a cell cycle
CC   inhibitor when expressed in a tumour cell. Modulating the expression of
CC   DMP1 can be used to treat tumours and other cancers. DMP1 can also be
CC   used for controlling expression of heterologous proteins. Antisense
CC   sequences and ribozymes can be used to inhibit expression of the
CC   transcription factor. Detecting the level and activity of DMP1 in cells
CC   is useful for detection of cancer cells or dysproliferative cells.
SQ   Sequence   2903 BP;    880 A;    662 C;    648 G;    713 U;


alignment_scores:
           Quality:   71.00            Length:    106
             Ratio:  1.109              Gaps:      5
 Percent Similarity:  60.377   Percent Identity: 25.472

alignment_block:
 US-09-174-937-8 x V15883   ..

 Align seq 1/1  to: V15883  from: 1  to: 2903

      17 ProProLeuIleLeuArgArgLeuLeuSerGluAspValGlyMetAspIl 33
         |||   :::|||:::::::  :::   :::||.|||:::      :::::
    1952 CCAAGAGUCAUCAUUCAGACGGUAGCUACAGAGGACAGCACUUCUUCAUU 2001

      33 eProPheGluGluGlyValLeuSerProSer......AlaAlaAspMetA 48
         :     ':'|:|   :::::::::   :::      ::::::::|:::
    2002 AUCCCAAGAGGAACUGACAGUUGAUAGUGAUCUUCAUUCAUCUGAUUUC. 2050

      48 rqProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArg 64
           |||||:|||:::::::|||:::        ||||   |||   :::
    2051 ..CCUGAGCCUCCAGAUGCACUAGAAGCA...GACACUUUCCCAGACGAA 2095

      65 IleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGl 81
         |||              :::|||:::   ||.................
    2096 AUUCCUCGGCCUAAGAUGACUAUACAACCAUCAUUUAAUAAUGCUCAUGU 2145

      81 ySerIleLeuSerAspAspAsn                 ..... .....L 89
          |||   :::|||||:::|||                          :
    2146 AUCUAAAUUCAGCGACCAAAAUAGCACAGAAC.GAUGAACAGUGUCAUGG 2195

      89 euAspSerProAspGluIle...AspIleAsnValAspGluLeuAspThr 104
         ::   :::   :::'|||'|   |'|   ::::::   :::   :::
    2196 CCAGAACAGAGGAAGAAAUUGCCGACACUGAC..UAAGCAGGAAGAACCG 2245

     105 ProAspGluAlaAspSer 110
         '|':::::::     |'|
    2246 CCGUCUGACUUAGCCAGU 2263
```

```
OM of: US-09-174-937-8 to  EST:*   Out_format : pfs

Date: Feb 10, 2000  1:33 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_SEQ/US09174937/runat_1_220100_095107_13273/app_query.fasta 1
 -DB=EST -QFMI=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -MINPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -STAPI=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pts -NORM=ext -MINLEN=0 -MAXLEN=10000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search Information block:
Query: US-09-174-937-8
Query length: 126
Database: EST:*
Database sequences: 4538634
Database length: 1887931982
Search time (sec): 1131.150000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation ..
gb_est5:D61567          -  451.00  958.15  1.3e-44   302  ! D61567 HUM418G12B Clontech huma
gb_est15:AA504063       -  439.00  933.52  3.1e-43   275  ! AA504063 nh40g09.s1 NCI_CGAP_Pr
gb_est12 AA296475       -  273.00  593.04  1.9e-24   235  ! AA296475 EST10971 Umbilical ve
gb_est32:AI722140       -  208.00  459.89  3.1e-13   604  ! AI722140 fd19g03.y1 Sugano Kawa
gb_gss1:FR00000132      -  182.50  376.96  3.1e-12   572  ! AL000418 F.rubripes BAC sequenc
gb_est1b:Z99377         -  135.00  277.58  1.1e-06   469  ! Z99377 HSZ99377 DKFZphamy1 Homo
gb_est14:AA396890       -  128.00  262.64  7.3e-06   467  ! AA396890 mr38g06.r1 Life Tech m
gb_est30:AI663516       -  128.00  260.70  9.4e-06   557  ! AI663516 uk33g11.y1 Sugano mous
gb_est38:AW044733       -  126.00  255.74  1.8e-05   593  ! AW044733 um13h05.y1 Sugano mous
gb_est39:AW106241       -  125.00  252.15  3.8e-05   677  ! AW106241 um25a08.y1 Sugano mous
gb_est9:AA096641        -  120.00  244.63  7.4e-05   507  ! AA096641 mo09f08.r1 Life Tech m
gb_est9:AA120357        -  120.00  246.06  8.7e-05   425  ! AA120357 mp89f04.r1 Soares 2NbM
gb_est11:AA210477       -  119.00  246.28  6.0e-05   359  ! AA210477 mu30h01.r1 Soares 2NbM
gb_est36:AI882015       +  118.50  241.01  0.0001    526  ! AI882015 ul30c05.y1 Sugano mous
gb_est36:AI931279       -  118.50  241.06  0.0001    524  ! AI931279 ul67b04.y1 Sugano mous
gb_gss14:AQ533091       -  118.00  239.16  0.0001    565  ! AQ533091 FPCI-11-351B1_TV FPCI-
gb_est36:AI876404       -  102.50  207.94  0.0081    474  ! AI876404 uk74f05.y1 Schiller mo
gb_gss10:AQ234674       +   92.50  188.72  0.0955    399  ! AQ234674 HS_2045_A2_F09_MR CIT
gb_gss5:AQ799604        -   87.50  175.39  0.5284    495  ! AQ799604 HS_4507_A2_F12_17A CIT
gb_gss5:AQ746111        -   84.50  169.93  2.33      791  ! AQ746111 HS_2275_A1_B04_MF CIT
gb_est10:AA134913       -   83.50  168.03  1.36      444  ! AA134913 zo26d03.s1 Stratagene
gb_est30:AI640514       -   81.50  162.17  2.99      513  ! AI640514 wa29a03.x1 NCI_CGAP_Pr
gb_est39:AW137099       -   81.50  161.62  3.09      539  ! AW137099 UI-H-BI1-acp-d-09-0-U1
gb_est38:AL043017       -   80.00  154.47  7.72      772  ! AL043017 DKFZp434C0322_r1 434 (
gb_est25:AI260912       -   79.00  157.27  5.68      591  ! AI260912 LP04928.5prime LP Dros
gb_est38:AW078345       -   79.00  155.06  7.16      602  ! AW078345 fe48c08.y1 Zebrafish W
gb_gss4:AQ669751        +   78.00  154.24  7.96      534  ! AQ669751 HS_5381_A1_H05_SP6E RP
gb_est39:AW052247       -   77.50  153.50  8.75      518  ! AW052247 rs33c10.y1 Sommer Pris
gb_est35:AL043881       -   77.00  152.52  9.92      514  ! AL043881 DKFZp434C2428_r1 434 (
gb_est35 AL046604       -   77.00  149.04  15.51     706  ! AL046604 DKFZp434L248_r1 434 (s
gb_est3:H01987          -   76.50  153.08  9.23      443  ! H01987 yj34g02.s1 Soares placen
gb_est17 AA599219       -   76.50  153.13  9.19      441  ! AA599219 ag34c10.s1 Jia bone mu
gb_est38:AL045474       +   76.50  154.20  8.00      400  ! AL045474 DKFZp434H185_r1 434 (s
gb_est10:AA160093       +   76.00  155.24  7.00      330  ! AA160093 zo75b03.r1 Stratagene
gb_est21:C90009         +   76.00  146.99  20.16     700  ! C90009 C90009 Dictyostelium dis
gb_est25:AI257069       -   76.00  150.13  11.45     526  ! AI257069 LP05329.5prime LP Dros
gb_gss9:AQ162744        +   76.00  147.46  19.00     671  ! AQ162744 mgxb0016M09r CUGI Rice
gb_gss13:AQ448951       +   76.00  148.78  16.04     595  ! AQ448951 mgxb0020020f CUGI Rice
gb_est10:AA193907       +   75.50  151.02  12.03     440  ! AA193907 rs06h11.r1 Sommer Pris
gb_est23:AI147356       -   75.50  150.75  12.46     451  ! AI147356 qg64a07.s1 Soares_test
gb_est30:AU057076       -   75.50  150.77  12.42     450  ! AU057076 AU057076 Oryza sativa
gb_est5:H76866          +   75.00  147.66  18.51     542  ! H76866 17297 Lambda-FLC2 Arabid
gb_est18:AA687506       -   75.00  149.33  11.83     469  ! AA687506 nf6g11.r1 NCI_CG
gb_est21:AA941201       -   75.00  144.52  25.71     722  ! AA941201 ol12c02.5prime LD
gb_est30:AB026258       -   75.00  142.72  34.86     557  ! AB026258 AB026258 Rattus n

seq_name: gb_est5:D61567

seq_documentation_block:
LOCUS       D61567              302 bp    mRNA             EST       14-DEC-1995
DEFINITION  HUM418G12B Clontech human fetal brain (lambda ZAP II) Homo
            sapiens cDNA clone GEN-418G12 5', mRNA sequence.
ACCESSION   D61567
VERSION     D61567.1  GI:970692
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 302)
  AUTHORS   Fujiwara T., Hirano H., Fukuda,T., Kuroda,K., Kitano,T., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shirada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takeda,T., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Fujiwara et al. (1995)
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:175481.
            Contact: Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
            Tel: 0886-65-2888
            Fax: 0886-37-1035
            Insert Length: 1881  Std Error: 1.11
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1..302
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GEN-418G12"
                     /clone_lib="Clontech human fetal brain polyA+ mRNA
                     (#6535)"
BASE COUNT       79 a     78 c     65 g     79 t     [illegible]
ORIGIN


alignment_scores:
            Quality:  451.00              Length:     [illegible]
              Ratio:  5.067                 Gaps:
 Percent Similarity:  98.889   Percent Identity:  98.889

alignment_block:
 US-09-174-937-8 x D61567  ..

 Align seq 1-[illegible]  to  [illegible]  from 1  to  302

       1 MetLeuLysSerCysSerArgAlaSerPheSerProSerValArgLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 ATGCTGAAGAGCTGTAGCCGTGCAT[illegible]CAGCCTCTGTTAGAAAGCCT 8
         [illegible]
      17 cProLeuIleLeuArgArgLeuL[illegible]AspVa.GlyMetArgLysP 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 TCCTCTCATCCTCAGAAGACTA[illegible]AGAGTAGGAATGATATCT 130
         [illegible]
      34 roPheGluGluGlyValLeuSerProSerAlaAlaAspMotArgProLeL 5
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 CCTTTGAAGAGGGGGTGCTCAGT[illegible]AGAGCAT[illegible]AGAATGAAGCT 194
         [illegible]
      51 ProProAsnSerLeuAspLeuAs.AspLysLysProArgArgIleLysLe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     195 CCTCCTAATTCTCTAGATCTTGA[illegible]AGATCAAAAGCCTCGAGGAATTAAGCT 254
         [illegible]
      67 uThrAlaProAsnIleAsnLeuSerLeuAspGluSerGluGlySerIleL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 CACAGCCCCAAATATCAACTCTT[illegible]AGTCAACAAGTCAAGGATTCAATCT 314
         [illegible]
```

```
      84 euSerAspAspAsnLeuAsp 90
         ||||||||||||||||||||
     281 TCTCTGATGATAACTTGGAC 300

seq_name: gb_est15:AA504063

seq_documentation_block:
LOCUS       AA504063       275 bp    mRNA            EST       20-AUG-1997
DEFINITION  nh40g03.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954881, mRNA
            sequence.
ACCESSION   AA504063
VERSION     AA504063.1  GI:2239030
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 275)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811027.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I M A G E Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 404  Std Error: 0.00
            Seq primer: -40m13 fwd  ET from Amersham
            High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1..275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE 954881"
                     /clone_lib="NCI_CGAP_Pr5"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from normal prostatic
                     epithelial cells, cDNA made by oligo-dT priming.
                     Non-directionally cloned.  Size-selected on agarose gel,
                     average insert size 600 bp.  "
BASE COUNT       80 a     66 c     62 g     67 t
ORIGIN


alignment_scores:
            Quality:  439.00            Length:     86
              Ratio:  5.105               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.837

alignment_block:
 US-09-174-937-8 x AA504063

 Align Seq 1/1  to: AA504063  from: 1  to: 275

      24 leuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValLe 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      18 CTACTGTCAGAGGATGTAGGAATGGACATCCCTTTGAAGAGGGCGTGCT 67

      40 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 57
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      68 GAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATC 117

      57 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 73
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     118 ITAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATATCAAT 167

      74 LeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAs 90
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     168 CTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTGGA 217

      90 pSerProAspGluIleAspIleAspValAspGluLeuAspThrProAspS 107
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 CAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCCGATG 267

     107 luAlaAsp 109
         |||||||
     268 AAGCAGAT 275

seq_name: gb_est12:AA296475

seq_documentation_block:
LOCUS       AA296475       236 bp    mRNA            EST       14-APR-1997
DEFINITION  EST10971 Umbilical vein endothelial cells II Homo sapiens cDNA 5'
            end, mRNA sequence.
ACCESSION   AA296475
VERSION     AA296475.1  GI:1948809
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Dougherty,B.A.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Morrin,M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellarino,S.M.,
            Phillips,C.A., Ryder,E.F., Scott,J.L., Shu,Jack,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Depeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,J., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hung,J.J., Li,H., Meissner,J.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl): 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Apr 14, 1993 this sequence version replaced gi:[illegible].
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..236
                     /organism="Homo sapiens"
                     /db_xref="AICC (inhost):194265"
                     /db_xref="taxon:9606"
                     /clone_lib="Umbilical vein endothelial cells II"
                     /cell_type="endothelial cell"
                     /dev_stage="fetus"
                     /note="Organ: umbilical vein; Vector: pBluescript SK-;
```

```
                    Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       66 a      54 c      57 g      55 t       4 others
ORIGIN


alignment_scores:
            Quality:  279.00              Length:     60
              Ratio:  4.982                 Gaps:      0
 Percent Similarity:  93.333   Percent Identity: 93.333

alignment_block:
 US-09-174-937-8 x AA296475   ..

 Align seg 1/1  to: AA296475  from: 1  to: 236

      24 LeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValLe 40
         ||||| |||||||||||||||| |||||||||| |||||||||||||||||
      55 CTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCT 104

      40 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 57
         ||||| |||||||||||||||||||||||||||||    ||||||      |
     105 GAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCNTAATTCTNTGGTTC 154

      57 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 73
         ||||| ||||||||||||||||||||||||||||||||||||||||| |||
     155 TTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATNTCAAT 204

      74 LeuSerLeuAspGlnSerGluGlySerIle 83
         ||||| |||||||||||||||||||||||||
     205 CTTTCTCTGGACTAAAGTGAAGGATCTATT 234

seq_name: gb_est32:AI722140

seq_documentation_block:
LOCUS       AI722140      604 bp    mRNA            EST       10-JUN-1999
DEFINITION  fd19g03.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2352052 5' similar to TR:O15073 O15073 KIAA0367 ;, mRNA sequence.
ACCESSION   AI722140
VERSION     AI722140.1  GI:5040469
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,I.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B.,
            Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948370.
            Other_ESTs: fd19g03.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2352052"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="[illegible] (phage resistant)"
                     /note="Vector: pT[illegible]; Site_1: [illegible] (GAATTCTTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [illegible]; 5' and 3'
                     double-stranded cDNA was ligated to a DraIII adaptor
                     (TGTTGGGCTA[illegible]) digested and gel purified to get DraIII
                     sites of the pME18S-FL3 vector (5' site TACTTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTTTG[illegible]AAAA[illegible] and 3' end
                     primer CGACTTGCAGTTCGAGCACA. "
BASE COUNT      184 a     145 c     159 g     121 t       [illegible] others
ORIGIN


alignment_scores:
            Quality:  208.00              Length:     [illegible]
              Ratio:  2.849                 Gaps:      [illegible]
 Percent Similarity:  72.277   Percent Identity: 44.5[illegible]

alignment_block:
 US-09-174-937-8 x AI722140   ..

 Align seg 1/1  to: AI722140  from: 1  to: 604

      26 SerGluAspValGlyMetAspIleProPheGluGluGlyValLeuSerPr 42
         ||   ||||||||||||||| |||  |||||  |||  |||||||||||
     126 TGTTGGCAAATGTGGGCATGGACCACATTTAT...CAAAGCAAGATGAGTTT 174

      42 uSerAlaAlaAspMetArgProGluProProAsnSerLeuAspLeu..... 57
         ||||||| |||   |||||| |  ||| |||||||||||||||||
     175 TGAAGGGGCAGAGAGTAGACCAGTTCACTCACTTGCTTCTAGAGG 222

      58 .....AsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIle 74
           ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
     223 GCCAAGGGGGTGTGTCTTCTCACAGCAAAAAAGTGTCGCTTCCCTTATT 272

      75 AsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLe 94
         ::|||| | |||| | | |||||                        ||
     273 AGTCTTTCATTGGACCAAAGTGAG..............GACGACTTTTT 307

      89 uAspSerProAspGluIleAspIleAsnValAspGlLeuAspThrProA 1[illegible]
         ||:|:::|:||:|:::::::|| |  |  |  |  |  |  |  |:::
     308 TGACACGCCAGATGACCTCGACATCAACTGTGATGACTGGACACTACCTTT 357

     106 spGluAlaAspSerPheGluTyrThrGlyA[illegible]sGlyLysSerLeuSerTrp 122
         |||||||||||  ||  |  |  |  |  |  |  ||||||
     358 ATGAAGCAGAT.........TATACAGACTAT......GAAATACACTTT 394

     123 Gln 123
         :::
     393 GAA 395

seq_name: gb_gss1:FR0009132

seq_documentation_block:
LOCUS       FR0009132     572 bp    DNA             GSS       16-SEP-1997
DEFINITION  [illegible]rubripes GSS sequence, clone CI HindIII [illegible], genomic sequence.
ACCESSION   AL000418
VERSION     AL000418.1  GI:2438273
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Trapnell,V.,
            Williams,G. and Brenner,S.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) MRC H. an Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB1 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Fugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 01 H20"
                     /clone="010H20aD1"
BASE COUNT      108 a    142 c    141 g    147 t     34 others
ORIGIN


alignment_scores:
           Quality:  182.50             Length:     74
             Ratio:  3.259                Gaps:      2
 Percent Similarity:  75.676   Percent Identity: 52.703

alignment_block:
 US-09-174-937-8 x FR0009132/rev    .

 Align seq 1/1  to: reverse of  FR0009132  from  1  to: 572

      51 ProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLe 67
         |||||:::::::||   |||:::          ::|||   |||||
     437 CCGCCCACCANTTTACCCCTGCAAGGAGGTCCAGGTCACCGTAAAAAGCT 388

      67 uThrAlaProAsnIleAsnLeuSerLeuAspGluSerGluGlySerIleL 84
         |:::|||||:::::||:::|||||||||||||' ||||||
     387 CTCCCCCCCTGGAATCAGCCTGTCGCTGGACCAGAGCGAG           348

      84 euSerAspAspAsnLeuAspSerProAspGlu''eAspIleAsnValAsp 100
            '|'||      :::::|||||||:::::|||||||'|||||||
     347 .....GACGACCTCGGGGAGACGCCAGACGACTCGACATCAACGTGGAT 303

     101 GluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGlyHisGl 117
         |||||||||||||||||||||:::|||   :::::|||||   |||
     302 GAGCTTGACACGCCGGACGAGGGGGACTACCTGGACTACACTGACCAC.  255

     117 yLysSerLeuSerTrpGlnGly 124
            :::   ::: |'::'|
     254 ....GAGACCGACTGGGAAGGT 237

seq_name: gb_est16:Z99377

seq_documentation_block:
LOCUS       Z99377          468 bp    mRNA            EST       22-SEP-1997
DEFINITION  HS299377 DKFZphamy1 Homo sapiens cDNA clone DKFZphamy1_1b2, mRNA
            sequence.
ACCESSION   Z99377
VERSION     Z99377.1  GI:2415617
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Korn B.  Wiemann S.  Ebert L.  and Poustka,A
  TITLE     Human ESTs (Korn,B. et al.)
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393564
            Contact: B. Korn
            Institution Molekulare Genomanalyse
            Deutsches Krebsforschungszentrum Heidelberg
            Im Neuenheimer Feld 506, D-69120 Heidelberg, FRG.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZphamy1_1b2"
                     /clone_lib="DKFZphamy1"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
BASE COUNT      155 a     84 c    114 g    115 t
ORIGIN


alignment_scores:
           Quality:  135.00             Length:     93
             Ratio:  2.077                Gaps:       6
 Percent Similarity:  69.892   Percent Identity: 59.785

alignment_block:
 US-09-174-937-8 x Z99377   ..

 Align seq 1/1  to: Z99377  from: 1  to: 468

      32 AspIleProPheGluGluGlyValLe'SerProSerAlaA'hAspM'tTr 79
         |||      ::|||     ::.:| ...   :::::   :::
       2 GATGATAGTATTGAAGCAGATATACTAGCTATAACTGGACCAGAG..... 46

      48 gProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgI 95
         ..    |||...|||||||:::::'|    :::       | :::
      47 ....GACCAGCCTGGCTCACTAGAAGTTAATGGAAATAAATTGAGAAAG. 91

      65 leLysLeuThrAlaProAsnIleAsnIleSerLeuAspGlnSerGluGly 81
         ||||||   '|||||:::|||:::'::::::|||||      '::::
      92 ..AAACTAATGGCTCCAGACATTAGCCTGACTCTGGATCCTAGTGATGGC 141

      82 SerIleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAs 98
         |||:::|||||||||||||   ||||||:::      ||||||||'::::
     140 TCTGTATTGTCAGATGAT...TTGGATGAAAGTGGGGAAATTGACTTA.. 184

      98 nValAspGluLeuAspThrProAspGln...AlaAspSerPheGluTyrT 114
            |||   | |:::|||||||:::'  ':::::::::   :  :::
     185 ....GATGGCTTAAACACACCGTCAGAGAATAGTAATGAGTTTGAGTGG. 229

     114 hrGlyHisGlyLysSerLeuSerTrpGln 123
                  :    ' ||||||:::
     230 ..........GAAGTCATTTCCTGGAAG 247

seq_name: gb_est14:AA396890

seq_documentation_block:
LOCUS       AA396890        467 bp    mRNA            EST       25-AUG-1997
DEFINITION  mr38g06.r1 Life Tech mouse embryo 15 5dpc 1067012 Mus musculus
            cDNA clone IMAGE 599770 5' similar to TR:O50844 O50844 NIP2.;,
            mRNA sequence.
ACCESSION   AA396890
VERSION     AA396890.1  GI:2049914
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria, Rodentia, Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Jerry,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stephens,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Leonard,S., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:365202
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1   467
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:599770"
                     /clone_lib="Life Tech mouse embryo 15 5dpc 10667012"
                     /tissue_type="embryo"
                     /dev_stage="15.5dpc embryos"
                     /lab_host="DH10B"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                     SalI; Site_2: NotI. Cloned unidirectionally.  Primer:
                     Oligo dT.  15.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT      143 a     92 c    140 g     91 t      1 others
ORIGIN


alignment_scores:
            Quality:  128.00             Length:     99
              Ratio:  2.098                Gaps:      5
 Percent Similarity:  68.539   Percent Identity: 38.202

alignment_block:
 US-09-174-937-8 x AA396890

 Align seq 1/1  to: AA396890  from: 1  to: 467

      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 43
         ||||||      :::   ...|||   ...........   ...   ..
      75 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAC 124

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :              |||:::     ||:::||||||.....|||
     125 ACTAGATGGAACTGATCCAGACAGACAGCCTCGCTCCTTAGAAGTTAATG 174

      59 spIhrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||.:  ||||||   |||||.  |||.:|||...
     175 GGAACAAAGTAAGGAAG...AAATTGATGGCCCTAGACATCAGCCTGACC 221

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||:  :::|||  |||...  |||'|:|||  ||||||...
     222 CTGGATCCTGGTGAAGACATTTGTGTGGGCCCATCAT  TTCGATGAAGC 268

      92 sAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
           |||:::|||.          :|||:|||||||:|||  :
     269 TGGAGAGGTTGACCTG......GAAGGCTTAGACACCCCATCAGAGAACA 312

     108 laAspSerPheGluTyr 113
         :::||:::|||||:::
     313 GTGATGATTTTGATTGT 329

seq_name: gb_est30:AI663516

seq_documentation_block.
LOCUS       AI663516      557 bp    mRNA            EST       10-MAY-1999
DEFINITION  uk33g11.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1970852 5' similar to TR:Q12982 Q12982 NIP2. ;, mRNA
            sequence.
ACCESSION   AI663516
VERSION     AI663516.1  GI:4767099
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota, Metazoa, Chordata, Craniata; Vertebrata, Mammalia,
            Eutheria, Rodentia, Sciurognathi, Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Shughtai,M., Theising,B., Ali,J., Yu,Y.,
            Person,B., Swaller,J., Ellington,M., Fedele,J., Harmon,N.,
            Ritter,E., Kohn,S., Shinn,T., Richardson,K., Taylor,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1-99
  JOURNAL   Unpublished (1999)
COMMENT     On May 19, 1999 this sequence version replaced gi:1477111.
            Other_ESTs: uk33g11.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:997592
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1971852"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACACAGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     (ATGTGGCCTT.....TTTTTTTTTT); 5'-1st-strand-cDNA was
                     ligated to a  DraIII adaptor (GCGCGCCACACTGTGTG); digested
                     and cloned into DraIII site of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACACAGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments of 5kb.  Library
                     constructed by  T. Sugano (The Tokyo Metropolitan
                     Institute of Medical Science).  Custom primers  for
                     sequencing.  5' end primer: CTTCACCATAAAAGTTCCC 3-1 3' end
                     primer CGACGTCGACGTCGAGGACA."
BASE COUNT      142 a     139 c     174 g     102 t
ORIGIN


alignment_scores:
            Quality:  128.00             Length:     99
              Ratio:  2.098                Gaps:      5
 Percent Similarity:  68.539   Percent Identity: 38.202

alignment_block:
 US-09-174-937-8 x AI663516

 Align seq 1/1  to: AI663516  from: 1  to: 557

      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 43
         ||||||      :::   ...|||   ...........   ...   ..
     279 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAC 328

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :              |||:::     ||:::||||||.....|||
     329 ACTAGATGGAACTGATCCAGACAGACAGCCTCGCTCCTTAGAAGTTAATG 378

      59 spIhrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||.:  ||||||   |||||.  |||.:|||...
     379 GGAACAAAGTAAGGAAG...AAATTGATGGCCCTAGACATCAGCCTGACC 425

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||:  :::|||  |||...  |||'|:|||  ||||||...
     426 CTGGATCCTGGTGAAGACATCTGTGTGGGCCCATCAT  TTCGATGAAGC 475
```

```
     92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
        ||:::|:|:::      :::   | :|  |||||:::|||   :
    473 TGGAGAGGTTGACCTG......GAAGGCTTAGATACCCCATCAGAGAACA 516

    108 laAspSerPheGluTyr 113
        ::|||::: |||||:::
    517 GTGATGAGTTTGAGTGG 533

seq_name: gb_est38:AW044733

seq_documentation_block:
LOCUS       AW044733      593 bp    mRNA            EST       18-SEP-1999
DEFINITION  um13h05.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2132217 5' similar to TR:O54940 O54940 NIP2L. ., mRNA
            sequence.
ACCESSION   AW044733
VERSION     AW044733.1  GI:5905330
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187044.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1004669
            Seq primer: custom primer used
            High quality sequence stop: 542.
FEATURES             Location/Qualifiers
     source          1..593
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2132217"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      135 a    169 c    182 g    107 t
ORIGIN


alignment_scores:
            Quality:  126.00            Length:     88
              Ratio:  2.100               Gaps:      5
 Percent Similarity:  68.182   Percent Identity:  38.636
```

```
alignment_block:
 US-09-174-937-8 x AW044733   ..

 Align seq 1/1  to: AW044733  from: 1  to: 593

     27 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSe 41
        ||||||      :::   :::|||   :::::::::::::   :::   ::
    341 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAC 39

     43 rAlaAlaAspMetArgProGlu   ProProAspSerLeuAspLeuAspLA 59
        :      |||:::    | :::|||||:::::::   :   :
    391 ACTAGATGGAACTGATCCAGACAGACAGCCTGGCTCCTTAGAAGTTAATG 44

     59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 77
        :::      |||...  ||||::  ||||||:::|||:::|| ||::
    441 GGAACAAAGTAAGGAAG...AAACTGATGGGCCCAGACATCAGCCTGACC 487

     76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
        ||||||   :::|||   |||:::   ||||||||||   ||||:::::
    488 CTGGATCCTGGTGAAGACTCTCTGTGTCCGATGAT...TTGGATGAAGC 534

     92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
        |||...||:::      .::   :|||||||||||::: |    :
    535 TGGAGAGGTTGACCTG...  GAAGGCTTAGATACCCCATCAGAGAACA 582

    108 laAspSerPheGlu 112
        ::|||:::||||||
    579 GTGATGAGTTTGAG 592

seq_name: gb_est39.AW106241

seq_documentation_block:
LOCUS       AW106241      677 bp    mRNA            EST       [illegible]
DEFINITION  um25a08.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:2225558 5' similar to SW:NIP2_MOUSE O54940 NIP2L/ADEN VIRUS
            E1B 19-KD PROTEIN-INTERACTING PROTEIN 2. [1] ., mRNA sequence.
ACCESSION   AW106241
VERSION     AW106241.1  GI:6076977
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1999 this sequence version replaced gi:[illegible].
            Other_ESTs: um25a08.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO [illegible], USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1005754
            Seq primer: custom primer used
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2225558"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
```

```
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an Oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTTT],
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT      193 a    159 c    210 g    124 t      1 others
ORIGIN


alignment_scores:
            Quality:  125.00              Length:     89
              Ratio:  2.083                Gaps:       5
 Percent Similarity:  67.416   Percent Identity:  59.202

alignment_block:
 US-09-174-937-8 x AW106241    ..

 Align seg 1/1  to: AW106241  from: 1  to: 677

      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 43
         ||||||      ...   ..:|||   ..........:   ...   ::
     249 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATGC 298

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :          |||.::     |||...||||||   ...|||
     299 ACTAGATGGAACTGATCCAGACAGACAGCCTGGTTCCTTAAAAGTTAATG 348

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
          ....   |||...   ||||||   ||| ||:..|||:.:|||:::
     349 GGAACAAAGTAAGGAAG   AAACTGATGGCCCCAGACATCAGCCTGACC 395

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   :::|||   |||:::   ||||| |||   ||||||:::
     396 CTGGATCCTGGTGAAGACTCTCTTGTGGTCCGATGAT...TTGGATGAAGC 442

      92 rAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
           |||:::'|:::      ...   |||| ||||||||:::|||   :
     443 TGGAGAGGTTGACCTG......GAAGGCTTAGATACCCCATCAGAGAACA 486

     108 laAspSerPheGluTyr 113
         ::|||:::|||||:::
     487 GTGATGAGTTTGAGTGG 503

seq_name. gb_est9.AA096641

seq_documentation_block:
LOCUS       AA096641      507 bp    mRNA            EST       25-OCT-1996
DEFINITION  mo09f08.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus
            cDNA clone IMAGE:553095 5' similar to TR:G558844 G558844 NIP2. ;,
            mRNA sequence.
ACCESSION   AA096641
VERSION     AA096641.1  GI:1642688
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Tholsing,E., Wylie,T., Lennon,J., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1642687.
            Contact:  Marra M,Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:333887
            Possible reversed clone; similarity to minus strand
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1..507
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:553095"
                     /clone_lib="Life Tech mouse embryo 10 5dpc 1 NIP 16"
                     /tissue_type="embryo"
                     /dev_stage="10.5dpc embryos"
                     /lab_host="DH10B"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally;  Primer:
                     Oligo dT.  10.5dpc embryos.  pCMV-SPORT2 vector."
BASE COUNT      133 a    119 c    160 g     95 t
ORIGIN


alignment_scores:
            Quality:  120.00              Length:     89
              Ratio:  2.000                Gaps:       5
 Percent Similarity:  67.416   Percent Identity:  57.303

alignment_block:
 US-09-174-937-8 x AA096641    ..

 Align seg 1/1  to: AA096641  from: 1  to: 507

      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 43
         ||||||      ...   ..:|||   ..........:   ...   ::
     229 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATGC 278

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :          |||.::     |||...||||||   ...|||
     279 ACTAGATGGAACTGATCCAGACAGACAGCCTGGTTCCTTAAAAGTTAATG 328

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
          ....   |||...   ||||||   ||| ||:..|||:.:|||:::
     329 GGAACAAAGTAAGGAAG...AAACTGATGGCCCCAGACATCAGCCTGACC 375

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   :::|||   |||:::   ||||| |||   ||||||:::
     376 CTGGATCCTGGTGAAGACTCTCTTGTGGTCCGATGAT...TTGGATGAAGC 422

      92 rAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
           |||:::'|:::      ...   |||| ||||||||:::|||   :
     423 TGGAGAGGTTGACCTG......GAAGGCTTAGATACCCCATCAGAGAACA 468

     108 laAspSerPheGluTyr 113
         ::|||:::|||||:::
     467 GTGATGAGTTTGAGTGG 483

seq_name. gb_est9.AA120357

seq_documentation_block:
LOCUS       AA120357      425 bp    mRNA            EST       17-FEB-1997
DEFINITION  mg89f04.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:579415 5'
            similar to TR:G558844 G558844 NIP2. ;, mRNA sequence.
ACCESSION   AA120357
```

```
VERSION     AA120357.1  GI:1677877
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Marra,M.  Hillier,L.; Allen,M.  Bowles,M.; Dietrich,N.; Dubuque,T.;
            Geisel,S.; Kucaba,T.; Lacy,M.  Le,M.  Martin,J.; Morris,M.;
            Schellenberg,K.; Steptoe,M.; Tan,F.; Underwood,K.; Moore,B.;
            Theising,B.; Wylie,T.; Lennon,G.; Soares,B.; Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 8, 1995 this sequence version replaced gi:800271.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:351063
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 212.
FEATURES             Location/Qualifiers
     source          1..425
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:576415"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      134 a     96 c    112 g     83 t
ORIGIN


alignment_scores
            Quality:  120.00             Length:     89
              Ratio:  2.000               Gaps:      5
 Percent Similarity:  67.416   Percent Identity: 37.079

alignment_block:
 US-09-174-937-8 x AA120357  ..

 Align seq 1/1  to: AA120357  from: 1  to:  425

      27 GluAspValGlyMetAspIleProPheGluGlyGlyValLeuSerProSe 43
         ||||||   :::   :..|||  :  :  :  :  :  :  :  :  :
       6 GAAGATTTTCCAATACCTTACCAGAAGAIGACAGCATTGAAGCAGATAC 55

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :          |||...     |||  ||||||:: ::::|||
      56 ACTAGATGGAACTGATCCAGACAGACAGCCTGGCTCCTTAGAAGTTAATG 105

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
            :::      |||...   ||||||   ||||||...|||...|||:::
     105 GGAACAAAGTAAGGAA3...AAACTGACCCCCCAGACATGACCCTGACC 152

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   :::|||   |||:::   |||   ||||||||||||:::
     153 CTGGATCCTGGTGAAGACTCTCTGTGGTCC...GATAATTTGGATGAAGC 199

      92 oAspGluIleAspIleAsnValAspGlnLeuAspThrProAspGlu...A 108
            |||:::|||:::      :::      ||||||||||::||||   :
     200 TGGAGAGGTTGACCTG.     GAAGGCTAAGACACCCCATCAGAGACCA 243

     108 laAspSerPheGluTyr 113
         ::|||:::||||||:::
     244 GTGATGAGTTTGAGTGG 260

seq_name: gb_est11:AA210477

seq_documentation_block:
LOCUS       AA210477      359 bp    mRNA            EST       19-FEB-1997
DEFINITION  mu30h01.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:640945 5'
            similar to TR:P55884 P55884 MIP2. , mRNA sequence.
ACCESSION   AA210477
VERSION     AA210477.1  GI:1807774
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Mammalia;
            Eutheria, Rodentia; Sciurognathi; Muridae, Murinae, Mus.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1406841.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:392937
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 55.
FEATURES             Location/Qualifiers
     source          1..359
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:640945"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      116 a     75 c     96 g     72 t
ORIGIN
```

```
alignment_scores:
            Quality:  119.00              Length:     89
              Ratio:  1.919                Gaps:       5
 Percent Similarity:  69.663   Percent Identity:  35.955

alignment_block:
 US-09-174-937-8 x AA210477

 Align seq 1/1  to: AA210477  from  1  to: 359

      27 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSe 43
         ||||||      :::   .  :::          ::::   :::   ::
       6 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAC 55

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :              |||:::      |||:::|||||||:::::|||
      56 ACTAGATGGAACTGATCCAGACAGACAGCCTGGCTCCTTAGAAGTTAATG 105

      59 spIhrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||:::   ||||||   ||'|||...|||...||...
     106 GGAACAAAGTAAGGAAG. .AAACTGAAGGCCCCAGACATCAGCCTGACC 152

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||.....|||   |||:  :  |||'|'|||   ||:::
     153 CTGGATTCTGGGTGAAGACTCTCTGTGGTCCGATGAT ..TTGGAATAAGC 199

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
            |||:....          ' ...|'|'|| ||||||:::|||   :
     200 TGGAGAGGTTGAACTG . . GAAAGCTTAGACACCCCATCAGAGAACA 243

     108 laAspSerPheGluTyr 113
         ::|||:::|||||:::
     244 GTGATGAGTTTGAGTGG 260

seq_name: gb_est36:AI882015

seq_documentation_block:
LOCUS       AI882015      526 bp    mRNA            EST       22-JUL-1999
DEFINITION  ul30c05.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2099816 5' similar to TR:O54940 O54940 NIP2L. ;, mRNA
            sequence.
ACCESSION   AI882015
VERSION     AI882015.1  GI:5567167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Thelsing,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1999 this sequence version replaced gi:1133619
            Other_ESTs: ul30c05.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St  Louis, MO 63109, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:996748
            Seq primer: custom primer used
            High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1..526
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2099816"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pBluescript SK(-); Site_1:
                     DraIII (CACTGTGTG); Site_2: DraIII (CACTATGTG); 5' strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TTTCTGGTCTATTGG]; digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTTGTG, 3' site CACTATGTG); XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <0.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer GTTTCTTCCTAAAAGTTCTCG and 3' end
                     primer CGACGTTGTAAAACGACGGCCA."
BASE COUNT      127 a    133 c    169 g     97 t
ORIGIN


alignment_scores:
            Quality:  118.00              Length:     91
              Ratio:  2.155                Gaps:       4
 Percent Similarity:  67.073   Percent Identity:  37.8 5

alignment_block:
 US-09-174-937-8 x AI882015   ..

 Align seq 1/1  to: AI882015  length: 1  to: 526

      27 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSe 43
         ||||||      :::   :::|  '  :::::::::::::   :::   ::
     282 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAC 331

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :              |||:::      |||:::|||||||:::::|||
     332 ACTAGATGGAACTGATCCAGACAGACAGCCTGGCTCCTTAGAAGTTAATG 381

      59 spIhrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||:::   ||||||   ||'|||...|||...||...
     382 GGAACAAAGTAAGGAAG...AAACTGAAGGCCCCAGACATCAGCCTGACC 428

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||.....|||   |||:  :  |||'|'|||   ||:::
     429 CTGGATCCTGGTGAAGACTCTCTGTGGTCCGATGAT...TTGGAATAAGC 475

      92 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 108
            |||:....          ' ...|'|'|| ||||||:::|||   :
     476 TGGAGAGGTTGAACTG......GAAAGCTTAGACACCCCATCAGAGAACA 525

seq_name: gb_est36:AI931278

seq_documentation_block:
LOCUS       AI931278      524 bp    mRNA            EST       22-JUL-1999
DEFINITION  ul67b04.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2135599 5' similar to TR:O54940 O54940 NIP2L. ;, mRNA
            sequence.
ACCESSION   AI931278
VERSION     AI931278.1  GI:5667769
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Thelsing,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
```

```
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948554.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1000275
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1..524
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2135594"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT      130 a    133 c    166 g     95 t
ORIGIN


alignment_scores:
            Quality:  118.50              Length:     82
              Ratio:  2.155                 Gaps:      4
 Percent Similarity:  67.073   Percent Identity: 37.805

alignment_block:
 US-09-174-937-8 x AI931278  ..

 Align seq 1/1  to: AI931278  from: 1  to: 524

      27 GluAspValGlyMetAspIleProPheGluGlyValLeuSerProSe 43
         ||||||      :::   ...|||   ...........  ...   ..
     282 GAAGATTTTCCAATACCTTTACCAGAAGATGATAGCATTGAAGCAGATAC 331

      43 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 59
         :              ||||::     |||...||||||.....:|||
     332 ACTAGATGGAACTGATCCAGACAGACAGCCTGGTTCCTTAGAAGTTAATG 381

      59 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 75
         :::      |||:::   ||||||   |||| |:::|||:::|||::: 
     382 GGAACAAAGTAAGGAAG   AAACTGATGGCCCAGACATCAGCCTGACC 428

      76 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 92
         ||||||   :::|||  ||||::  ||||| ||  |||||||:::
     429 CTGGATCCTGGTGAAGACTCTCTGTGGTCCGATGAT...TTGGATGAAGC 475

      92 sAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu 107
          |||:::|||:::      :::   |||||| |||||||:::|||
     476 TGGAGAGGTTGACCTG      GAAGGCTTAGATACCCCATCAGAG 515
```

SmithWaterman 4.1x
Copyright (c) 1993 - 2003 Compugen Ltd.

Protein - protein search, using sw model

Run on: February 19, 2003, 14:15:36 ; Search time 16 Seconds
(without alignments)
1404.631 Million cell updates/sec

Title: PCT-US02-32032-1
Perfect score: 283
Sequence: 1 MKIGILGVVALVIGSWIHLTIS.....GGQQGVQQQQGIVGPLISAGPA 254

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 75134422 residues

Word size: 6

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Query No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB	ID	Description
1	283	100.0	254	1	PCT-US02-32032-1
2	88	31.1	2	A62066	phosphocarrier pr
3	85	30.0	124	2	A37951	CapA - Sweetsop (
4	84	29.7	139	2	A26892	Mspa box protein
5	84	29.7	219	2	I39671	RPE15 protein - ra
6	84	29.7	251	2	I22839	hypothetical pr***
7	84	29.7	327	2	JC5462	gamma-filadin pro**
8	84	29.7	337	2	A5 918	P35Parilde protein
9	84	29.7	456	2	S01992	glutenin low molec
10	84	29.7	459	2	I49982	glutenin low molec
11	83	29.3	372	2	I52621	mitogen-activated
12	83	29.3	474	2	T05923	glutenin low molec
13	83	29.3	481	2	T34333	hypothetical pr***
14	83	29.3	545	2	I40912	ribonucleoprotein
15	83	29.3	696	2	I24576	hypothetical prote
16	83	29.3	939	2	I23752	hypothetical prote
17	82	29.0	114	2	S60190	gliadin - Zea mays
18	82	29.0	453	2	S63383	POP2 protein - yea
19	82	29.0	465	2	B40256	protein F21B7.6 [i
20	82	29.0	471	2	I23750	hypothetical pr***
21	82	29.0	475	2	I24751	hypothetical prote
22	82	29.0	492	2	S19147	EMF-2 protein - Ar
23	82	29.0	494	2	S62402	ferumin 2 precurso
24	82	29.0	499	2	S05893	hypothetical pr***
25	82	29.0	504	2	S01942	prumin 2 precursor
26	82	29.0	517	2	S28229	diacylglycer a kin
27	82	29.0	519	2	S00661	ferumin - Welwitsc
28	82	29.0	522	2	B42804	immunodominant typ
29	82	29.0	531	2	A03444	hypothetical pr***

A;Reference number: A82035; MUID:20406833; PMID:1095401
A;Accession: A82035
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-92 <HEI>
A;Cross-references: GB:AA000052; GB:AB000052; NID:g557104; PIDN:AAB57471.1; GSPDB:GN00024
A;Experimental source: serogroup 7b; strain N1966; biotype El Tor
C;Genetics:
A;Gene: vc.055
A;Map position: 1

Query Match 5.0%; Score 8; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 ISAEGSQA 89
||||||||
Db 66 ISAEGSQA 73

RESULT 3
A37941
hairin - Swedish turnip (fragments)
C;Species: Brassica napus var. rapifera (Swedish turnip, rutabaga)
C;Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 12-Apr-1996
C;Accession: A37941
R;Svendsen, I.; Nicolaou, G.; Gosbey, I.; Genoz, N.
Carlsberg Res. Commun. 54, 241-249, 1989
A;Title: Isolation and characterization of a trypsin inhibitor from the seeds of kohlrab
A;Reference number: A37941; MUID:91282906; PMID:2490369
A;Accession: A37941
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-124 <SVE>
C;Superfamily: wheat alpha-amylase inhibitor

Query Match 5.0%; Score 8; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 QQQQQQQQ 255
||||||||
Db 86 QQQQQQQQ 93

RESULT 4
A26892
Mopa box protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 05-Nov-1996
C;Accession: A26892
R;Dubeaux, C.; Baculin, M.; Galliot, B.; Mehler, E.
Mol. Cell. Biol. 7, 2099-2106, 1987
A;Title: DNA sequences homologous to the Drosophila opa repeat are present in murine mRN
A;Reference number: A26892; MUID:87257766; PMID:2885744
A;Accession: A26892
A;Molecule type: mRNA
A;Residues: 1-139 <DUB>
A;Cross-references: GB:M16564; NID:g200142; PIDN:AAA39860.1; PID:g78759

Query Match 5.0%; Score 8; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 QQQQQQQQ 251
||||||||
Db 1 QQQQQQQQ 11

RESULT 5
I22819
EPEI% protein - alfalfa (fragment)
C;Species: Medicago sativa (alfalfa)
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 23-Jul-1996
C;Accession: I22819
R;Xia, X.; Du, S.; Erickson, L.
Genome 39, 9-16, 1996
A;Title: A moderately repetitive DNA sequence in alfalfa is [illegible]
A;Reference number: Z19621; MUID:96208794
A;Accession: I22819
A;Status: preliminary; translated from EMBL/[illegible]
A;Molecule type: DNA
A;Residues: 1-219 <XIA>
A;Cross-references: EMBL:U90656; NID:g1[illegible]; PIDN:AAB[illegible]; PID:g1[illegible]
A;Experimental source: leaf
C;Genetics:
A;Gene: EPEI%

Query Match 5.0%; Score 8; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 QQQQQQQQ 252
||||||||
Db 161 QQQQQQQQ 168

RESULT 6
I22819
hypothetical protein F57A8.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change [illegible]
C;Accession: I22819
R;McMurray, A.
submitted to the EMBL Data Library, April 1997
A;Reference number: Z19621
A;Accession: I22819
A;Status: preliminary; translated from EMBL/[illegible]
A;Molecule type: DNA
A;Residues: 1-251 <WIL>
A;Cross-references: EMBL:Z70781; PIDN:CAA94831.1; GSPDB:GN00024; [illegible]
A;Experimental source: clone F57A8
C;Genetics:
A;Gene: CESP:F57A8.2
A;Map position: 5
A;Introns: 41/1; 85/1; 117/3

Query Match 5.0%; Score 8; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 QQQQQQQQ 251
||||||||
Db 71 QQQQQQQQ 78

RESULT 7
JS0402
gamma-zeathin precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change [illegible]
C;Accession: JS0402
R;Sudiyama, T.; Fatalski, A.; Scott, D.
Plant Sci. 44, 205-209, 1986
A;Title: The nucleotide sequence of a wheat gamma-zeathin gene
A;Reference number: JS0402
A;Accession: JS0402
A;Molecule type: DNA
A;Residues: 1-427 <SUD>
A;Experimental source: cv. Yamhill
C;Comment: Gamma-zeathins belong to a subclass of the 27 kDa gamma-zeathins
C;Superfamily: zeathin
C;Keywords: storage protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-327/Product: gamma-zeathin #status predicted <MAT>

```
  Query Match                3.0%;  Score 8;  DB 2;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     245 QQQQQQQG 252
           ||||||||
Db     247 QQQQQQQG 254


RESULT 8
A60918
FMRFamide precursor - fruit fly (Drosophila virilis)
C;Species: Drosophila virilis
C;Date: 31-Dec-1993 #sequence_revision 03-Feb-1994 #text_change 17-Mar-1999
C;Accession: A60918
R;Taghert, P.H.; Schneider, L.E.
J. Neurosci. 10, 1929-1942, 1990
A;Title: Interspecific comparison of a Drosophila gene encoding FMRFamide-related neurop
A;Reference number: A60918; MUID:90285722; PMID:2113087
A;Accession: A60918
A;Molecule type: DNA
A;Residues: 1-339 <TAG>
A;Cross-references: GB:M32612
C;Genetics:
A;Gene: FlyBase:Dvir/Fmrf
A;Cross-references: FlyBase:FBgn0013078
C;Superfamily: FMRFamide polyprotein
C;Keywords: amidated carboxyl end; neuropeptide; tandem repeat
F;1-17/Domain: signal sequence #status predicted <SIG>
F;114-122/Product: FMRFamide-related peptide V1 #status predicted <FV1>
F;153-161/Product: FMRFamide-containing peptide V2a #status predicted <FV2A>
F;164-172/Product: FMRFamide-containing peptide V2b #status predicted <FV2B>
F;175-182/Product: FMRFamide-containing peptide V2c #status predicted <FV2C>
F;185-193/Product: FMRFamide-containing peptide V2d #status predicted <FV2D>
F;196-202/Product: FMRFamide-containing peptide V4 #status predicted <FV4>
F;214-222/Product: FMRFamide-containing peptide V5 #status predicted <FV5>
F;225-231/Product: rMkFamide-containing peptide V6 #status predicted <FV6>
F;234-242/Product: FMRFamide-related peptide V7 #status predicted <FV7>
F;245-252/Product: FMRFamide containing peptide V8 #status predicted <FV8>
F;122/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;161/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;172/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;182/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;193/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;202/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;222/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;241/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl
F;242/Modified site: amidated carboxyl end (Asn) (amide in mature form from following gl
F;252/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gl

  Query Match                3.0%;  Score 8;  DB 2;  Length 339;
  Best Local Similarity   100.0%;  Pred. No. 6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     245 QQQQQQQG 252
           ||||||||
Db     301 QQQQQQQG 308


RESULT 9
S01992
glutenin, low molecular weight chain precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 20-Aug-1999
C;Accession: S01992
R;Pitts, E.G.; Kafalski, J.A.; Hedgcoth, C.
Nucleic Acids Res. 16, 11376, 1988
A;Title: Nucleotide sequence and encoded amino acid sequence of a genomic gene region fo
A;Reference number: S01992; MUID:89083577; PMID:3205747
A;Accession: S01992
A;Molecule type: DNA
A;Residues: 1-356 <PIT>
A;Cross-references: EMBL:X07747; NID:g21702, PIDN:CAA30570.1; PID:g21783
```

```
C;Superfamily: [illegible]

  Query Match                [illegible]
  Best Local Similarity   [illegible]
  Matches    [illegible]

Qy     245 [illegible] 252
           [illegible]
Db     265 [illegible] 272


RESULT 10
[illegible]
[illegible]
C;Species: [illegible]
C;Date: [illegible]
C;Accession: [illegible]
R;Anders[illegible]
submitted to the EMBL Data Library, [illegible]
A;Reference number: [illegible]
A;Accession: [illegible]
A;Status: [illegible]
A;Molecule type: mRNA
A;Residues: 1-[illegible] <ANE>
A;Cross-references: EMBL:[illegible]; NID:[illegible]; PIDN:AAB[illegible]; [illegible]
A;Experimental source: [illegible]
C;Genetics:
A;Map position: [illegible]
C;Function:
A;Description: Seed storage protein in the [illegible] genus
C;Superfamily: [illegible]
C;Keywords: seed; storage protein

  Query Match                [illegible]
  Best Local Similarity   [illegible]
  Matches    [illegible]

Qy     245 [illegible] 252
           [illegible]
Db     261 [illegible] 268


RESULT 11
I52621
mitogen-activated protein kinase MAPK [imported] - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-[illegible]-[illegible] #sequence_revision [illegible] #text_change [illegible]
C;Accession: I52621
R;Jonak, C.; Kiegerl, S.; Ligterink, W.; Barker, P.J.; Huskisson, N.S.; Hirt, H.
Gene 229, 171-181, 1999
A;Title: Characterization of a novel plant MAP kinase gene encoding MMK2
A;Reference number: Z41147; MUID:[illegible]; PMID:[illegible]
A;Accession: I52621
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-[illegible] <JON>
A;Cross-references: EMBL:[illegible]; NID:[illegible]; PIDN:CAA[illegible]; [illegible]
C;Superfamily: [illegible]

  Query Match                [illegible]
  Best Local Similarity   [illegible]
  Matches    [illegible]

Qy     244 [illegible] 251
           [illegible]
Db     417 [illegible] 424


RESULT 12
I05926
glutenin [illegible]
C;Species: Triticum aestivum (common wheat)
C;Date: 30-Apr-[illegible] #sequence_revision [illegible] #text_change [illegible]
```

C;Accession: I05923
R;Maruyama, N.; Ichise, K.; Fa' J. I F. ; .' T Kusaka S.; Matsumura, Y.; Take
Eur. J. Biochem. 255, 739-745, 1998
A;Title: Identification of major wheat allergens by means of the Escherichia coli expres
A;Reference number: Z15459; MUID:98409296; PMID:9738916
A;Accession: I05923
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-374 <MAR>
A;Cross-references: EMBL:AB008497, NID:g2582837; PIDN:BAA23162.1; PID:g2582838
A;Experimental source: cultivar hard red spring ICW
C;Superfamily: gliadin
C;Keywords: seed; storage protein

  Query Match 3.0%; Score 8; DB 2; Length 374;
  Best Local Similarity 100.0%; Pred. No. 6.5;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 GGGGGGGG 252
        ||||||||
Db 276 GGGGGGGG 283

RESULT 13
I34333
hypothetical protein K06A1.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: I34333
R;Fulton, L.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid K06A1.
A;Reference number: Z21507
A;Accession: I34333
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-381 <FUL>
A;Cross-references: EMBL:U23449; PIDN:AAC24296.1; GSPDB:GN00020; CESP:K06A1.1
A;Experimental source: strain Bristol N2; clone K06A1
C;Genetics:
A;Gene: CESP:K06A1.1
A;Map position: 2
A;Introns: 112/1, 214/2, 271/2, 314/2

  Query Match 3.0%; Score 8; DB 2; Length 381;
  Best Local Similarity 100.0%; Pred. No. 6.6;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 GGGGGGGT 253
        ||||||||
Db 112 GGGGGGGT 119

RESULT 14
I00912
ribonucleoprotein homolog F21B7.26 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
C;Accession: I00912
R;Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con
colosis, A.; Ecker, J.R.
submitted to the EMBL Data Library, January 1998
A;Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.
A;Reference number: Z14208
A;Accession: I00912
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-395 <SHI>
A;Cross-references: EMBL:AC002560; NID:g2618677; PID:g2809257
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 1
A;Introns: 54/2, 81/3, 99/2, 152/3, 176/3, 222/2, 252/2, 344/1

C;Superfamily: unassigned ribonucleoprotein repeat containing protein
  Query Match 3.0%; Score 8; DB 2; Length 395;
  Best Local Similarity 100.0%; Pred. No. 6.8;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 GGGGGGGG 251
        ||||||||
Db 376 GGGGGGGG 383

RESULT 15
T24576
hypothetical protein T06D8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24576
R;Palmer, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19909
A;Accession: T24576
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-396 <WIL>
A;Cross-references: EMBL:Z49130; PIDN:CAA89066.1; GSPDB:GN00022; CESP:T06D8.3
A;Experimental source: clone T06D8
C;Genetics:
A;Gene: CESP:T06D8.3
A;Map position: 2
A;Introns: 112/2, 169/2, 221/2, 264/2, 314/2

  Query Match 3.0%; Score 8; DB 2; Length 396;
  Best Local Similarity 100.0%; Pred. No. 6.8;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 GGGGGGGG 252
        ||||||||
Db 381 GGGGGGGG 388

Search completed: February 19, 2003, 13:17:50
Job time : 21 secs

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2000, 19.33.05 , Search time 3770.17 Seconds
(without alignments)
967.408 Million cell updates/sec

Title: US-09-174-937-6
Perfect score: 966
Sequence: 1 atggaggaggagacagagtt.....tgaagctgaaagaaaagcct 966

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831992 residues

Database : EST:*

Word size : 0

Number of hits that pass the thresnold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8.*
28: gb_est9.*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8.*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	269.2	27.9	275	34	AA[illegible]	AA[illegible] nm40g09.s
2	220.4	22.9	286	31	AA[illegible]	AA[illegible] EST10971
3	204.2	21.1	302	24	[illegible]	[illegible] HUM418G12B
4	173	17.9	312	20	[illegible]	[illegible] HUMEA02P Atr
5	168.2	17.4	423	20	[illegible]	[illegible] EST41767 Hu
6	166.2	17.2	369	20	[illegible]	[illegible] EST73534 Fe
7	163.8	17.0	449	29	AA[illegible]	AA[illegible] nq75d03.r

	8	156	15.5	499	45	AI437189 fb38gc06.y AI437189 fb38gc06.y
	9	156.2	15.2	359	33	AA459821 vg55e03.r AA459821 vg55e03.r
	10	154	15.9	516	40	AA914577 ve02e07.r AA914577 ve02e07.r
	11	139.6	14.5	452	28	AA067188 mm38b12.r AA067188 mm38b12.r
	12	139.6	14.5	443	28	AA068579 mm45e06.r AA068579 mm45e06.r
	13	135	14.0	382	28	AA113114 zn69h02.r AA113114 zn69h02.r
c	14	124.2	12.9	526	43	AI180114 EST223951 AI180114 EST223951
	15	123.2	12.9	441	21	T94657 T94657 ye37e08.r1
	16	119.8	12.4	281	20	T31983 T31983 EST41923 Hu
c	17	117.2	12.1	496	48	AI620433 tu47b05.x AI620433 tu47b05 x
	18	112	11.6	394	29	AA129380 zn84e05.r AA129380 zn84e05.r
	19	111.4	11.5	559	79	FR0014613 AL005822 F.rubripe
	20	107.2	11.1	604	51	AI722140 fd19g03.y AI722140 fd19g03.y
	21	103.2	10.7	467	31	AA280204 zt05c05.r AA280204 zt05c05.r
	22	102.2	10.6	357	41	AI006991 ua78a04.r AI006991 ua78a04.r
c	23	91.8	9.5	518	100	AQ318295 RPCI11-10 AQ318295 RPCI11-10
c	24	91.8	9.5	546	100	AQ318293 RPCI11-10 AQ318293 RPCI11-10
	25	90.8	9.4	397	43	AI202887 qr31b01.x AI202887 qr31b01.x
c	26	89.2	9.2	469	87	AQ769043 HS_3191_A AQ769043 HS_3191_A
c	27	85.2	8.8	551	29	AA129317 zn84e05.s AA129317 zn84e05.s
c	28	83.6	8.7	388	89	AQ892241 HS_2191_A AQ892241 HS_2191_A
	29	80	8.3	452	21	T68254 T68254 ye40g01.r1 T68254 ye40g01.r1
	30	78.8	8.2	239	79	FR0014614 AL005858 F.rubripe
	31	71.2	7.4	212	72	AV293206 AV293206 AV293206
	32	70	7.2	186	20	T32180 T32180 EST44614 Hu
	33	64	6.6	577	69	AW106241 AW106241 um25a08.y AW106241 um25a08.y
	34	61.2	6.3	577	64	AW052742 AW052742 fi14021B10 AW052742 fi14021B10
	35	60.6	6.2	519	62	AI976981 AI976981 EST271498
	36	60.4	6.2	542	40	C91038 C91038 C91038 Diet
	37	49.6	5.1	478	40	AA980800 ua45g06.r AA980800 ua45g06.r
c	38	49.6	5.1	516	49	AI642182 uc45h08.x AI642182 uc45h08.x
	39	47	4.9	467	33	AA396890 mr38g06.r AA396890 mr38g06.r
c	40	47	4.9	572	79	FR0022128 AL000418 F.rubripe
	41	45.4	4.7	562	51	AI727670 AI727670 BNLGH1855
	42	45.2	4.7	236	79	FR0000218 AL000440 F.rubripe
	43	44.4	4.6	468	35	Z99377 Z99377 HSZ99377 DK
	44	44.4	4.6	705	47	AI517653 AI517653 GH28696.5
c	45	44	4.6	619	69	AI978259 fi14042E06 AI978259 fi14042E06

ALIGNMENTS

RESULT 1
AA504063
LOCUS AA504063 275 bp mRNA EST 20-AUG-1997
DEFINITION nh40g09.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954880, mRNA sequence.
ACCESSION AA504063
VERSION AA504063.1 GI:2239030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 275)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:811027.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Ea Id B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 404 Std Error: 0.00
Seq primer: -40m13 fwd FT from Amersham
High quality sequence stop: 219
FEATURES Location/Qualifiers
source 1..275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954880"
/clone_lib="NCI_CGAP_Pr5"
/sex="male"
/tissue_type "prostate"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from normal prostatic epithelial cells; cDNA made by oligo dT priming; Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."
BASE COUNT 80 a 66 c 62 g 67 t
ORIGIN

Query Match 27.8%; Score 269.2; DB 94; Length 275;
Best Local Similarity 98.9%; Pred. No. 2e-67;
Matches 270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 42 tcaagacgaaatggactactgt[illegible] 102
Db 3 TCGACACGAAATGGACTACTGTCAGAGAATCTACGAATCTACACTCTTTTCAATTAT 62

Qy 103 gtgctgagtcccagtgctgcagacatcagtctgaagcctcattatttctttaat 162
Db 63 GTGTTGAGTCCCAGTGTTGCAGACATCAGTCTGAACCTCCTAATTTCTTTAAT 122

Qy 163 gacactcatctcagagaaatcaaa[illegible] 222
Db 123 GACACTCATCCTCGGAGAATCAAGTCACAGCCCCAAATATCAATCTTTCTCTCGACTAA 182

Qy 223 agtgaaggatctattctctctgatcatcaacttggacagtccaaatgaaatgaatcatcat 282
Db 183 AGTGAAGGATCTATTCTCTCTGATCATAATTTGGACAGTCCAGATCAAATCACATCAAT 242

Qy 283 gtggatgaaacttgatacccccgatgaagcagat 315
Db 243 GTGGATGAACTTGATAGCCCCGATGAAGCAGAT 275

RESULT 2
AA296475
LOCUS AA296475 236 bp mRNA EST 14-APR-1997
DEFINITION EST10971 Umbilical vein endothelial cells II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA296475
VERSION AA296475.1 GI:1948809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 236)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Ma,M.D., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., File,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.,Jr., Kelley,J.M., Kelley,T.P., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Norton,A.J.,Jr., Olsvik,O., Oswald,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,L.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Yu,G.-L., Bednarik,D.P., Cao,L., Chapela,M.A., Coleman,T.A., Collins,F.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,K.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,T., Holden,E., Kim,A.K., Korak,D.L., Kunsch,C., Marsden,I., Li,H., Melassas,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.-L., Fu,K.-M.,

```
            Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693143.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..236
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):194265"
                     /db_xref="taxon:9606"
                     /clone_lib="Umbilical vein endothelial cells II"
                     /cell_type="endothelial cell"
                     /dev_stage="fetus"
                     /note="Organ: umbilical vein; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       66 a     54 c     57 g     55 t      4 others
ORIGIN


  Query Match           23.3%;  Score 233.4;  DB 31;  Length 236;
  Best Local Similarity   97.0%;  Pred. No. 1.9e-56;
  Matches  231;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 gaggaggagacagaagttccttgagctcggaaccaggatatcaagaccaaaatggactactg 63
          ||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAGGAGGAGACAGAGTTCCNTTGAGCTCGGAACCAGGATATCAAGACCAAATGGACTACTG 60

Qy      64 tcagaggatgtaggaatggacatccccttgaagagggcgtgctgagtcccagtgctgca 123
           ||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db      61 TCAGAGGATGTAGGAATGGACATCCCCTTGAAGAGGGCGTGCTGAGTCCCAGTGCTGCA 120

Qy     124 gacatgaggcctgaacctcataattctctgatcttaatgacactcatcctcggagaatc 183
           ||| |||||||| |||| |||| |||       |||||||| |||||||||||||||||
Db     121 GACATGAGGCCTGAACCTCNTAATTCTNTGGTTCTTAATGACACTCATCCTCGGAGAATC 180

Qy     184 aagctcacagcccaaatatcaatctttctctggaccaaagtgaaggatctattct 239
           ||| |||| |||||| | |||||||||     |||||||||||||||||||||
Db     181 AAGTTCACAGCCCAAATNTCAATCTTTCTC.GGACCAAAGTGAAGGATCTATTCT 236


RESULT   3
D61567
LOCUS       D61567        302 bp    mRNA            EST       14-DEC-1995
DEFINITION  HUM418G12P Clontech human fetal brain polyA+ mRNA (#6535) Homo
            sapiens cDNA clone GEN-418G12 5', mRNA sequence.
ACCESSION   D61567
VERSION     D61567.1  GI:970692
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 302)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Fujiwara et al. (1995)
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785681.
            Contact: Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co., Ltd.
            463-10 Kagasuno Kawauchi-cho, Tokushima, 771-01 Japan
            Tel: 0886-65-2888
            Fax: 0886-37-1035
            Insert Length: 1881  Std Error: 0.
            High quality sequence stop: 302.
FEATURES             Location/Qualifiers
     source          1..302
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GEN-418G12"
                     /clone_lib="Clontech human fetal brain polyA+ mRNA
                     (#6535)"
BASE COUNT       79 a     78 c     65 g     79 t      1 others
ORIGIN


  Query Match           21.1%;  Score 211.2;  DB 24;  Length 312;
  Best Local Similarity   99.5%;  Pred. No. 8.2e-49;
  Matches  204;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      56 gactactgtcagaggatgtaggaatggacatccccttgaagagggcgtgctgagtccca 115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTGAAGAGGGCGTGCTGAGTCCCA 157

Qy     116 gtgctgcagacatgaggcctgaacctcataattctctgatcttaatgacactcatcctcg 175
           |||||||||||||||||||| |||| |||||||| |||||||||||||||||||||||||
Db     158 GTGCTGCAGACATGAGGCCTGAACCTCNTAATTCTNTGGTTCTTAATGACACTCATCCTCG 217

Qy     176 gagaatcaagctcacagcccaaatatcaatctttctctggaccaaagtgaaggatctattct 235
           |||||||||| |||||| | |||||||||     |||||||||||||||||||||||||||
Db     218 GAGAATCAAGTTCACAGCCCAAATNTCAATCTTTCTC.GGACCAAAGTGAAGGATCTATTCT 277

Qy     236 ttctctctgatgataacttggagaag 261
           ||| |||| |||||| | |||||||||
Db     278 TTCTCTCTGATGATAACTTGGACAAG 302


RESULT   4
Z32816
LOCUS       Z32816        312 bp    mRNA            EST       8-MAY-1994
DEFINITION  HHEA02P Atrium cDNA library Human heart Homo sapiens mRNA clone
            HEA02P, mRNA sequence.
ACCESSION   Z32816
VERSION     Z32816.1  GI:483499
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Auffray,C., Behar,G., Bohn,M., Bouchier,C., Brahimi,K.,
            Derignes,M.D., Dupraz,P., Hilbert,P., Imbert,P., Kahn,M.N.,
            Lorenzo,F., Mitchell,G., Marilet-Hams,L.F., Pinchotti,D.,
            Sebastiani-Kabakchie,S. and Tessier,A.
  TITLE     IMAGE, molecular interrogation of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III Sci. Vie 317 (2), 175-179 (1994)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:[illegible].
            Contact: Genzentrum Muenchen
            Laboratorium fuer molekularbiologie
            Am Klopferspitz 19a, 8033 Martinsried, Germany
            Email: dbmailer@vms.klinik.uni-mainz.de
            single read.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEA 2P"
                     /clone_lib="Human cDNA library, H. sapiens"
```

```
BASE COUNT       97 a     55 c     59 g    101 t
ORIGIN


  Query Match             17.9%;  Score 173;  DB 20;  Length 312;
  Best Local Similarity   72.5%;  Pred. No. 9.7e-40;
  Matches  224;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     544 atcattgtgtttgcrgcctgttttctgcccgaccagcagtcgggcggattaccactatgtc 603
           ||  ||||||||| | ||||||  |||| || || |||| | |  | ||    ||  |
Db       1 ATTTTTGTGTTTGCTGTCTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAGATACCTG 60

Qy     604 atggaaaatcttttcctatatgtaataagtactttagagttgatggtagctgaagactat 663
           ||||| |||||||   ||||| ||  | || || ||| |  | ||||| ||| ||||
Db      61 ATGGACAATCTTTTTAAATAIGTTATTGGCACTTTGGAGCTACTAGTAGCAGAAAACTAC 120

Qy     664 atgattgtgtacttgaatggtgcaaccccaaaaaggaggatgccagggctaggctggatg 723
           ||||| || || || ||||||||||  .  ||| |  |||||  | || || ||| |
Db     121 ATGATATTTTATTTAAATGGTGTAACAATCCAAAGAAAAATGCCCAGTCTGGGATGGTTC 180

Qy     724 aagaaatgctaccagatgattgacagacggtttgaggaagaatttgaaatcattcatcatt 783
           | | |||| || |||   ||||| ||| ||||  || | ||| | |||||  | ||||||
Db     181 AGGIAATGTIATCAGCAAATIGATAGAAGGTACGGGAAAATCTAAAATCCCTAATCATT 240

Qy     784 attcatccatcttcgttcatcagaacaatcctgctgtgaccgcgacctttataaagttca 843
           |  |||  ||||||| ||| |||  |  |||||| ||| |||| |||||  || ||
Db     241 GTACATCCTTCTTGGTTTATCAGAACACCTCTGGCTGTTACAAGACCATTTATTAGCTCG 300

Qy     844 aaatttcagc 853
           ||||||||||
Db     301 AAATTCAGC 309


RESULT   5
T31973
LOCUS       T31973         423 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST41767 Human Uterus Homo sapiens cDNA 5' end similar to similar
            to rhoGAP protein, mRNA sequence.
ACCESSION   T31973
VERSION     T31973.1  GI:614071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,P., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,F.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.P., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC14305
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699405b
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..423
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):102311"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Uterus"
                     /note="Organ: uterus"
BASE COUNT      128 a     71 c     90 g    131 t      3 others
ORIGIN


  Query Match             17.4%;  Score 168.2;  DB 20;  Length 423;
  Best Local Similarity   70.2%;  Pred. No. 2.6e-38;
  Matches  252;  Conservative    3;  Mismatches  105;  Indels    2;  Gaps    2;

Qy     512 attcaggatactatgggacggtctcaatgccatcattgtgtttgccgctattctctc 571
           ||   ||||| ||| |||| ||  ||||||||||  ||||||||  | || ||
Db      33 ATGGGGGATATTATNGGGATGGATTAAATGCCATTGTTGTGTTTGCTGTCTGTTTCATG 92

Qy     572 cagacaggcagtcgggcgggattaccactatgtcatggaaaatcttttcctatatgtaatag 631
           | || || |||| | |  | ||    ||  | |||||  ||||||  |  |  | |
Db      93 CTGAAAGTAGTCAGCCTAACTATAGATACCTGATGGACAATCTTTTTAAATATGTTATTG 152

Qy     632 gtactttagagttgatggtagctgaagactatatgattgtgtacttgaaatcctgcaatcc 691
           | |||||  ||| |  | |||||  ||| |||| |||| || || || ||||||||||
Db     153 GCACTTTGGAGCTATTAGTAGCAGAAAACTACATGATAGTTTATTTAAATGGTGTAACAA 212

Qy     692 caagaaggaggatgccagggctaggctggatgaagaaatgctaccagatgattgacagacg 751
           | |||| |  |||||  | || |  |  |  ||  |.  |  |  |  |  |  |  |
Db     213 CTCCAAAGAAAAATGCCCAGTCTGGGATGGTTCAGGAAATCATTATCAGCAAATTGATAGA 272

Qy     752 ggtt-gaggaagaatttgaaatcattcatccatcttcgttcatcagaacaatcctgctgtg 810
           ||||  ||||  ||| | ||||     |||||||| |||||  |||| | | | | | | |
Db     273 GGTTACGGGAAAAATCTAAAATCCCTAATCATTGTACATCCTTCTTGGTTTATCAGAACAC 332

Qy     811 atccttgctgtgac-acgacctttttataaagttcaaaattcagcagtaaaattaaatatg 868
           | || |||| |  | | |||||||| || ||  ||||     |||||  | | |
Db     333 CTTCTGGCTGTTANAAGGCCCTTTTATTAGCTCGGAATTTCAGCCAAAAATTTGTTATT 391


RESULT   6
T05645
LOCUS       T05645         368 bp    mRNA            EST       3-JUN-1993
DEFINITION  EST03534 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDD71, mRNA sequence.
ACCESSION   T05645
VERSION     T05645.1  GI:316795
KEYWORDS    EST
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Homo sapiens"
```

```
                     /db_xref="ATCC (inhost):82344"
                     /db_xref="taxon:9606"
                     /clone="HFBDD71"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT       88 a    106 c     90 g     80 t      4 others
ORIGIN


  Query Match             17.2%;  Score 100.2;  DB 20;  Length 368;
  Best Local Similarity   68.9%;  Pred. No. 9.6e-38;
  Matches  239;  Conservative    0;  Mismatches  107;  Indels    1;  Gaps    1;

Qy     589 gattaccactatgtcatggaaaatcttttcctatatgtaataagtactttagagttgatg 648
           || ||||||||  |||||||| || || ||||| || || || || |  |||||| |  ||
Db       2 GACTACCACTACATCATGGAGAACCTCTTCCTGTACGTNATCAGCAGCTTAGAGCTCCTG 61

Qy     649 gtagctgaagactatatgattgtgtacttgaatggtgcaacccccaagaaggaggatgccca 708
           || ||||| |||| |||| || ||| ||  ||||| || ||  |  ||||||||||
Db      62 GTGGCTGAGGACTACATGATCGTNTACCTGANCGGTGCCACGCCCCGGCGGAGGATGCCT 121

Qy     709 gggctaggctggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttg 768
           ||  | |||||| |||||||| ||||||||||||| |||  |  |||| |||| ||  ||
Db     122 GGAATCGGCTGGTTGAAGAAGTGCTACCAGATGATCGACCG-GAGTTGCGGAAAAACCTG 180

Qy     769 aaatcattcatcattgttcatccatcttgttcatcagaacaatccttgctgtgacacga 828
           || || || ||||| || || || || |||||||||  | ||  | || ||  |  | ||
Db     181 AAGTCCTTGATCATCGTCCACCCCTCGTCGTTCATTCGGACTGTGGTGGCCATCTTTGGC 240

Qy     829 cctttataagttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaactc 888
           |||'| || |'   || |||| ||  || |'  | || ||  | ||||   ||| ||
Db     241 CCTTCATCAGCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCTTGGAAGACCTN 300

Qy     889 agtggggctgatcccaatggattgcatccaacattccagagagcatcat 935
           ||  |||| |||||   | |||| || ||||   | || |
Db     301 GAGCAACTCATCCCTATGGAACAGGTCCAGATCCCAGACTTCGTCCT 347


RESULT   7
AA137873
LOCUS       AA137873      449 bp    mRNA            EST       09-FEB-1997
DEFINITION  mq75d03 r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:584549 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA137873
VERSION     AA137873.1  GI:1700244
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394228.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:359197
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 4[illegible].
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:584549"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="STLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT  From MC 5[illegible], a highly tumorigenic derivative of
                     the K-1735 (mouse) melanoma. Average Insert Size: 1.0 kb;
                     Vmi:ZAP XR Vector; 5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAG[illegible]TT 3'"
BASE COUNT      147 a     76 c     124 g     102 t
ORIGIN


  Query Match             17.0%;  Score 163.8;  DB 2;  Length 449;
  Best Local Similarity   65.2%;  Pred. No. 5.1e-[illegible];
  Matches  276;  Conservative    0;  Mismatches  137;  Indels    1;  Gaps    2;

Qy     337 gatcccatagcaagaaggatttcgcccgtgtttcgactctattccagaatatacccgcc 396
           |  |||| | ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      36 GAGCCCAAACCCAAAACTACTGAACTCATTACTAAAGGTTCGATTACTGAGTACACAGTT 95

Qy     397 gaagaggaagcggaaggacaacccgcttcgatgcgtggtcattggaaaacaaaggcag 456
           |'| ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      96 ACAGAAGAAAAGGGCGATGGACGTTGTTCGCTAATGTTCAGTATTGGAGACACAGTACTAT 155

Qy     457 cgcattgacatgaaggtcatcgtgcctcacgaggagctcatt tcaatcaggaggattca 516
           |  | |||||||  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     156 AGGGTGGACATGAAGGCGATCGTTCCCTACAAAAAGCTTATTATCAATCCTT  . .  . 206

Qy     517 ggatattatgggacggctctaaatgccattattgttgtttcttttttttcccaga 575
           |||| ||'|||||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     207 GGATATTATGGGGACGGCTTAAATGCTATTGTTGTTATTTGCTTTCTTTCCATAATCTTA 266

Qy     576 cagcagtcgggcggattaccactatatcatgaaaatctttttcatatatctaaaaactac 635
           |'  |||  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     267 GAGTGGTCAGCTAATTATAATGATTTTTATGTTCATTTCTTTTTAATATGTTTTTTGTAT 326

Qy     636 tttagagttgatggtagcctgaagactatatgattgtgtacttgaatggtgcaacccccaag 695
           |||'||| |' |||||  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db     327 TTTAGAGCTTATTGGTATCAGAATTATTAATTTTCATTATTAAATTGTTGTTATTTTTG 386

Qy     696 aaggaggatgccagggctaggctggatgaagaaatgctaccagatgattgacagacggttt 755
           |||  |  ||||  |  |  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db     387 AAGAAAAATGCCCAGTTTGGGATCGTTATCGTTATGCTACCAGTTAATTTTTTTTTTTAAT 446

Qy     756 gag 758
           |||
Db     447 AAG 449


RESULT   8
AI437189
LOCUS       AI437189      499 bp    mRNA            EST       06-MAR-1999
DEFINITION  fb28g06.y1 Zebrafish [illegible] 3' [illegible] 5', mRNA sequence.
ACCESSION   AI437189
VERSION     AI437189.1  GI:4288571
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,S., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
```

```
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,F.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138723.
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF'"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     (5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3'];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      163 a    110 c    117 g    109 t
ORIGIN


  Query Match             16.6%;  Score 160;  DB 46;  Length 499;
  Best Local Similarity   64.7%;  Pred. No. 6.7e-36;
  Matches  236;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy    587 cggattaccactatgtcatgaaaaatcttttctatatgtaataagtactttagagttga 646
          ||||||||| ||  | ||||||||||| || || |||||  |||| | ||| ||| ||
Db      4 CAGATTACCATTACCTGATGGAAAATCTTCTTCTGTATGTTGTAAGCAGTTTGGAGATGC 63

Qy    647 tggtagctgaagactatatgattgtgtatttcatgggtgcaacccaaagaaggaggatgc 706
          || | |.|| | ;|  |||| | || |  |||| | ||| ||  | || |  ||||
Db     64 TGGTGGCGACATTATCCTGATCATCTACATTCATGGAGGAACGGCGGCGGAGTAAAATGC 123

Qy    707 cagggctaggctggatgaagaaatgctaccaaatgattgacagacggttgaggaagaatt 766
          |  || | |||| |||| ||  |||| |||||| |  |||| || ||
Db    124 CCGGCATTAGCTGGCTCAAGAAGTGCTACCAGATGATCGACAGAAGACTGAGAAAAAACC 183

Qy    767 tgaaatcattcatcattgttcatccatcttgattcatcagaacaatccttgctgtgacac 826
          | ||||  | ||||  |||||| || |||| ||  |  ||  || | ||  |  |
Db    184 TAAAATCTCTGATCATCATCATCCCTCTGCTTATTCCGACGGTCATCGCCGATTTTCA 243

Qy    827 gaccttttataagttcaaaatt 848
          ||| || || ||   || |||| | ||| ||  || || | |||| |  ||| |
Db    244 AACCCTTCATCAGCGTGAAGTTCATCAATAAGATCCGTTATGTCTACAGTTTTGAATAGT 303

Qy    887 tcagtgggctgatcccaatgga 908
          |      | |  | || |||||  || || || |||||  | | |
Db    304 TGGAAAAGTTCGTTCCCATGGATCATATTCATATCCCAGAATGCATCTTACAATATAAAG 363

Qy    947 tgaagctg 954
          | | | |
Db    364 AGGAACGG 371


RESULT   9
AA450831
LOCUS       AA450831      368 bp    mRNA            EST       04-JUN-1997
DEFINITION  vg55e03.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865276 5' similar to TR:G558844 G558844 NIP2., mRNA
            sequence.
ACCESSION   AA450831
VERSION     AA450831.1  GI:2164501
KEYWORDS    EST
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Scarpa,P., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1344744.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509364
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /strain="C57BL6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:865276"
                     /clone_lib="Beddington mouse embryonic region"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally; Primer:
                     Oligo dT. Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings. Excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 kb to 3.0 kb).
                     Referenced in Development 121, 2479-2488 (1995)"
BASE COUNT      110 a     66 c     98 g     94 t
ORIGIN


  Query Match             16.2%;  Score 156.2;  DB 33;  Length 368;
  Best Local Similarity   67.4%;  Pred. No. 7.8e-35;
  Matches  254;  Conservative    0;  Mismatches  123;  Indels    1;  Gaps    2;
```

```
Qy    376 tctattccagaatatacggccgaagaggaaccggaggacaaccggctttggatgacagtg 435
          || ||| | || || || ||| ||| ||  | ||   || | |||   |  |
Db      1 TCGAIIACIGAGIACACAGCCACAGAAGAAAAGGGCGATGGACGCCGCTGG-CCAIGTIC 59

Qy    436 gtcattggagaacaagagcagcgcattgacatgaaggtcatcgagccctacaggagagtc 495
          |||||||| || || ||  |  | ||||||||||  ||||||||||| |  |||
Db     60 AGGATTGGAGAGCAGGACIACAGGGTGGACAIGAAGGCGATCGAGCCCTACAAAAAAGTI 119

Qy    496 atttctcaccggaggagattcaggatactatgggacggtctaaatgccatcattgtgttt 555
          |||   || |||||| ||         ||||||||||  ||||||||||  |||| |||
Db    120 ATTAGCCATGGAGGATAT·········IAIGGGACGGCTIAAATGCCAITGTIGTATTI 170

Qy    556 gccgcctgtttttctgccagacagcagtcgggcggattaccactatgtcatggaaaatctt 615
          || | ||||||  |||| || ||  ||| ·  | ||   ||| | ||||| |||||
Db    171 GCTGTCTGTTICATGCCIGAGAGTGGTCAGCCTAACTATAGATATTTGATGGACAATCTC 230

Qy    616 ttcctatatgtaataagtactttagagttgatggtagctgaagactatatgattgtgtac 675
          ||   |||||| ||  | |'|||||||| || ||||||| ||  |||| |||||  | |||
Db    231 TTTAAATATGTTATTGGCACTTTAGAGCIGIIGGTAGCAGAGAACTACATGATCATTTAC 290

Qy    676 ttgaatggtgcaaccccaagaaggaggatgccagggctaggctggatgaagaaatgctac 735
          || |||||||| ||  |  |||| |  |||||  |  | || ||| | | || |||||||
Db    291 TTAAATGGTGCGACAACICGAAGAAAAATGCCCAGTTTGGGATGGCTCAGGAGATGCTAC 350

Qy    736 cagatgattgacagacg 752
          |||   ||'|| |||||
Db    351 CAGCAAATTGATAGACG 367


RESULT  10
AA914577
LOCUS       AA914577      516 bp    mRNA            EST       14-APR-1998
DEFINITION  vz02e07.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1314564 5' similar to TR:Q12982 Q12982 NIP2. ;, mRNA
            sequence.
ACCESSION   AA914577
VERSION     AA914577.1  GI:3053969
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407399.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:684860
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1..516
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1314564"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="1 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pIT2(r-Bam
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Description: mRNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTAGCAATCTTAATTGGAGGCCGCGGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adapters (Pharmacia); cDNA was digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pIT2 vector.
                     RNA provided by Dr. Mikita Ko. Mouse 8-day whole Library
                     constructed and normalized by Benito Soares and M.Fatima
                     Bonaldo."
BASE COUNT      148 a    110 c    118 g     14 t
ORIGIN


  Query Match           15.9%;  Score 154;  DB 4;  Length 516;
  Best Local Similarity   69.0%;  Pred. No. 3.7e-14;
  Matches  211;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy    659 actatatgattgtgtacttgaatggtgcaaccccaagaaggaggatgccagggctaggct 718
          |||| |'|||  | |||| | |  |  |  |  |  |  |  |  |  |  |  |  |
Db      3 ACTACATGATCATTTACTTAAATGGTGGGACAACTTGAAGAAAAATGCCCAGTTTGGGAT 62

Qy    719 ggatgaagaaatgctaccagatgattgacagacggttgaaaaaaaaattttgaaatgattga 778
          || | | || ||||||||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     63 GGCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGCAAAAAACCTTAATTTCTTAA 122

Qy    779 tcattgttcatccatcttggttcatcagaaacaatcctttgctgttacacgacgacctttataa 838
          ||'|||| '| |' ||||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    123 TCATTGTGCACGGCICTTGGTTTAICCCAACAICTTTGTGGCTGTTACACGACGATTTIAIAA 182

Qy    839 gttcaaaattcagcagtgaaaattaaattttt..atgattttatcaggaattcaatgaaatgaa 898
          |'|'|'| ||||'   || | |  |  |  |  |  |  |  |  |  |  |  |  |
Db    183 GTTCAAAGTTCAGICACAAGATTACATAICTTCTTTAATTTGGCIAAGTTCCCAACTCTTC 242

Qy    899 tcccaatggattgcatccacattccagaaagcatcatcaaatatt.aattttgaaatgaaaa 958
          .. |' |||||  |  |  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db    243 TTCCCATGGAGTATGICGGCATACCAGAGTTGCATCAAACAA.TTGATCAACATTTAATG 302

Qy    959 aaaagc 964
          |||'|
Db    303 GAAAAC 308


RESULT  11
AA067188
LOCUS       AA067188      452 bp    mRNA            EST       6-FEB-1997
DEFINITION  mm38b12.r1 Stratagene mouse melanoma (#937910) Mus musculus cDNA
            clone IMAGE:503775 5' similar to TR:O08844 O08844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA067188
VERSION     AA067188.1  GI:1564954
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407485.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```

```
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:317623
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:523775"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      145 a     90 c     108 g     109 t
ORIGIN


  Query Match             14.5%;  Score 139.6;  DB 28;  Length 452;
  Best Local Similarity   69.3%;  Pred. No. 5.5e-30;
  Matches  190;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     671 tgtacttgaatggtgcaaccccaagaaggaggatgccagggctaggctggatgaagaaat 730
           |||||| |||||||| ||  |  |||| |  |||||  |  | || ||| | | || ||
Db       2 TTTACTTAAATGGTGCGACAACTCCAAGAAAAATGCCCAGTTTGGGATGGCTCAGGAGAT 61

Qy     731 gctaccagatgattgacagacggttgaggaaagaatttgaaatcattcatcattgttcatc 790
           ||||||||   ||||| ||||| || |||||  ||  | || ||  | ||||||| || |
Db      62 GCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTAAGTCTCTAATCATTGTGCACC 121

Qy     791 catcttggttcatcagaacaatcctttgctgtgacagacgtttttataagttcaaaattca 850
           | |||||||| ||| ||||| | || ||||| |||||||| ||||| |||||||| ||||
Db     122 CCTCTTGGTTTATCCGAACACTTCTGGCTGTTACACGACCATTTATCAGTTCAAAGTTCA 181

Qy     851 gcagtaaaattaaatatgtcaatagcttatcagaaactcagtgggctgatcccaatggatt 910
           |   ,| || | ||:  || || |  | || |    | |||  | || ||||| |
Db     182 GTCAGAAGATCAGATACGTCTTTAACTTGGCCGAGCTGGCAGAGCTTGTTCCCATGGAGT 241

Qy     911 gcatccacattccagagagcatcatcaaatattga 944
           |  || ||| ||||| ||||||    |    |
Db     242 ATGTCGGCATACCAGAGTGCATCAAACAGTATCA 275


RESULT  12
AA068579
LOCUS       AA068579      443 bp    mRNA            EST       05-FEB-1997
DEFINITION  mm45e06.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:524492 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA068579
VERSION     AA068579.1  GI:1575939
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802463

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318330
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:524492"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      143 a     89 c     106 g     105 t
ORIGIN


  Query Match             14.5%;  Score 139.6;  DB 28;  Length 443;
  Best Local Similarity   69.3%;  Pred. No. 5.5e-30;
  Matches  190;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     671 tgtacttgaatggtgcaaccccaagaaggaggatgccagggctaggctggatgaagaaat 730
           |||||| |||||||| ||  |  |||| |  |||||  |  | || ||| | | || ||
Db       2 TTTACTTAAATGGTGCGACAACTCCAAGAAAAATGCCCAGTTTGGGATGGCTCAGGAGAT 61

Qy     731 gctaccagatgattgacagacggttgaggaaagaatttgaaatcattcatcattgttcatc 790
           ||||||||   ||||| ||||| || |||||  ||  | || ||  | ||||||| || |
Db      62 GCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTAAGTCTCTAATCATTGTGCACC 121

Qy     791 catcttggttcatcagaacaatcctttgctgtgacagacgtttttataagttcaaaattca 850
           | |||||||| ||| ||||| | || ||||| |||||||| ||||| |||||||| ||||
Db     122 CCTCTTGGTTTATCCGAACACTTCTGGCTGTTACACGACCATTTATCAGTTCAAAGTTCA 181

Qy     851 gcagtaaaattaaatatgtcaatagcttatcagaaactcagtgggctgatcccaatggatt 910
           |   || || | |||  || |  | || |    | |||  | || ||||| |
Db     182 GTCAGAAGATCAGATACGTCTTTAACTTGGCCGAGCTGGCAGAGCTTGTTCCCATGGAGT 241

Qy     911 gcatccacattccagagagcatcatcaaatattga 944
           |  || ||| ||||| ||||||    |    |
Db     242 ATGTCGGCATACCAGAGTGCATCAAACAGTATCA 275


RESULT  13
AA113114
LOCUS       AA113114      382 bp    mRNA            EST       11-AUG-1997
DEFINITION  zn68h02.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            IMAGE 563379 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA113114
VERSION     AA113114.1  GI:1664694
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
```

```
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
 COMMENT    On May 5, 1995 this sequence version replaced gi:797955.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2034  Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1..382
                     /organism="Homo sapiens"
                     /db_xref="GDB:4593190"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:563379"
                     /clone_lib="Stratagene HeLa cell s3 937216"
                     /sex="female"
                     /dev_stage="HeLa S3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. HeLa S3
                     epithelioid carcinoma cells grown to semi-confluency
                     without induction. Average insert size: 1.5 kb; Uni-ZAP XR
                     Vector. ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      131 a     59 c     93 g     99 t
ORIGIN
```

```
  Query Match             14.0%;  Score 195;  DB 28;  Length 382;
  Best Local Similarity   65.2%;  Pred. No. 1.1e-28;
  Matches  234;  Conservative    0;  Mismatches  115;  Indels   10;  Gaps    2;

Qy     372 agagtctattcaagaatatacggccgaaggaggaacgggaggacaaccggctttggatgac 431
           || || ||| | ||||| || || | |||||  || ||    || | |||    |
Db      31 AGGCICAAIIACTGAATACACAGCAGCAGACCAAAAAGAAGATGGACGACGCTGGCGIAG 90

Qy     432 ag-tggtcattggagaacaagagcagcgcattgacatgaaggtcatcgagccctacagga 490
           || |  ||||||||||||| || ||  |  ||| |||||||  || || |||||  |  |
Db      91 TGTTCAGGATTGGAGAACAGGACCACAGGGI GATATGAAGGCAATTGAACCCTATAAAA 150

Qy     491 gagtcatttctcacggaggagattcaggatactatggggacggtctaaatgccatcattg 550
           ||| ||   || || | | ||          |||||||  ||  ||||||||||  |||
Db     151 AAGITATCAGCCAIGGGGGAIAT---------TATGGGGATGGATTAAATGCCATTGTIG 201

Qy     551 tgtttgccgcctgttttctgccagacagcagtcgggcggattaccactatgtcatggaaa 610
           ||||||| | ||||||  |||| || || |||| | |  | ||    ||  | |||||  |
Db     202 TGTTTGGTGTCIGTTICATGCCIGAAAGIAGICAGCCTAACTATAGATACCTGATGGACA 261

Qy     611 atcttttcctatatgtaataagtactttagagttgatggtagctgaagactatatgattg 670
           |||||||   ||||| ||  | ||||| || |  | ||||| ||| |||| ||||| |
Db     262 ATCTTTTTAAATATGTTAITGGCACITIGGAGCTATTAGTAGCAGAAAACTACATGATAG 321

Qy     671 tgtacttgaatggtgcaacccccaagaaggaggatgccagggctaggctggatgaagaaa 729
           | || || |||||||||||  |  |||| ||||||  | || ||   |    || ||
Db     322 TTTATTTAAATGGTGCAACAACTCGAAGAAAAATGCCCAGTCTGGGAATGGCTCAGGAA 380
```

```
RESULT  14
AI180114/c
LOCUS       AI180114      526 bp    mRNA            EST       20-JAN-1999
DEFINITION  EST223851 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
            RSPCQ31 3' end, mRNA sequence.
ACCESSION   AI180114
VERSION     AI180114.1  GI:3730752
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mayer,M., Lieskovsky,I.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
 COMMENT    Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..526
                     /organism="Rattus sp."
                     /db_xref="ATCC (dbEST):1096411"
                     /db_xref="taxon:10114"
                     /clone="RSPCQ31"
                     /clone_lib="Normalized rat spleen, Bento Soares"
                     /note="Organ: spleen; Vector: pBluescript SK-; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      151 a     108 c     108 g     154 t
ORIGIN
```

```
  Query Match             12.9%;  Score 124.2;  DB 4;  Length 526;
  Best Local Similarity   69.7%;  Pred. No. 1.7e-28;
  Matches  168;  Conservative    0;  Mismatches   73;  Indels    7;  Gaps    0;

Qy     724 aagaaatgctaccagatgattgcaagaaggttagtaagaaattgaaatcattcatgatt 783
           | || ||||||||||   ||  |||  | | | |  |  |  |  |  |  |  |||
Db     525 AGGAGATGCTACCAGCAAATCAA AGA AAT AAGAAAAAGCTAAAGTCTCTAATCATT 466

Qy     784 gttcatccatcttggttcatcagaacaatgcttgcatgtgacaaaaaattttatagattta 843
           |  |  | |  |||||||| |||          |  |  |  |  |  |  |  |  |
Db     465 GTGCACCCCTCTTGCTTTATCCGAACACTTCTCCCTCTATAAAAATAATTATTATCAGTTCA 406

Qy     844 aaattcagaagtaaaatt aaatatttttaattttttaaaa taaatttttatgtccca 903
           |||||||  | | |||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db     405 AAATTCAGTCAGAAGATTAGATAGTTTTAATTTTGCTAAATTTTTTTTAGTTTGTTTTC 346

Qy     904 atggattgcatcaaatttccagaaagcatcatcaaatattgaattgaagctgaaaaaaaaa 963
           |||||  | | |||  ||| |||  |  |  |  |  |  |  |  |  |  |  |
Db     345 ATGGAGTACGTTGGGATAGCAGATTGCTTCAAGATGTTTTTTTTTTTTTTTTTTTTTTTC 286

Qy     964 c 964
           |
Db     285 G 285
```

```
RESULT  15
T94657
LOCUS       T94657         441 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye37e08.r1 Stratagene [illegible] (#937216), [illegible] 5', mRNA cDNA clone
            IMAGE:119942 5', mRNA sequence.
ACCESSION   T94657
VERSION     T94657.1  GI:728145
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
```

```
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 28,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1632
            High quality sequence stops: 282 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1632  Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /db_xref="GDB:488231"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:119942"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT, normal lung.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      137 a     70 c     95 g    131 t      8 others
ORIGIN


  Query Match           12.9%;  Score 123.2;  DB 21;  Length 441;
  Best Local Similarity   70.7%;  Pred. No. 3.2e-25;
  Matches  164;  Conservative    0;  Mismatches   68;  Indels    0,  Gaps    0;

Qy    713 tatgctggatgaagaaatgctaccagattattacagacggttgaggaagaatttgaaat 772
          | || ||| | | |||||| || |||   ||| || ||| |||| |||| ||| | ||||
Db      1 TGGGATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAAAAATCTAAAAT 60

Qy    773 cattcatcattgttcatccatcttggttcatcagaacaatccttgctgtgacacgaccctt 832
          | | |||||||| |||| |||||||| ||  |||| | || ||||| ||| |||| |
Db     61 CCCIAAICAIIGIACAICCIICIIGGTIIAICAGAACACIICIGGCIGIIACAAGACCAI 120

Qy    833 ttataagttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaactcagtg 892
          ||||  | || ||||||||  ||||||||  | ||   ||  ||  |||||||    |
Db    121 TTATTAGCIGGAAAIICAGCCAAAAAAIIAGATACGTGIITAATTIGGCAGAACIAGCAG 180

Qy    893 ggctgatcccaatggattgcatccacattccaagagcatcatcaatattga 944
          ||  ||  |||| |||| | | | |  | ||  ||  |  |||| |   |  ||||
Db    181 AACTTGTCCCCATGGAATACGTTGGCATACCAGAATGCATAAAACAAGTTGA 232


Search completed: February  7, 2000, 19:33:09
Job time: 6699 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2000, 20:29:12 ; Search time 284.47 Seconds
(without alignments)
2705.350 Million cell updates/sec

Title: US-09-174-937-1
Perfect score: 3076
Sequence: 1 ccacgcgtccgcccacgcgt.....ggcatggtggtgcatgcctg 3076

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	331.6	10.8	384	1	Q59527	Human brain Expres
c 2	301.8	9.8	3095	1	V19075	Homo sapiens NGAL
3	300.6	9.8	3099	1	V99912	Human sialoadhesin
4	300.6	9.8	3099	1	V99911	Human sialoadhesin
5	295.8	9.6	45546	1	X23520	Human kidney amino
c 6	293.2	9.5	53577	1	T18551	Human polycystic k
c 7	293.2	9.5	53577	1	T94106	Human PKD1 locus b
c 8	293.2	9.5	53526	1	T94101	Human PKD1 gene. H
c 9	293	9.5	14796	1	V27941	Survivin gene. Mod
c 10	293	9.5	7146	1	V38933	Nucleic acid seque
c 11	286.6	9.3	2930	1	V34997	Human chemokine pr
c 12	285.6	9.3	1913	1	T76515	DNA encoding macro
c 13	285.6	9.3	2923	1	T99233	Coding sequence fo
c 14	285.6	9.3	2923	1	X32817	Macrophage derived
15	285.2	9.3	10475	1	N80616	Genomic clone enco
16	285.2	9.3	10475	1	N80643	Sequence of human
c 17	285	9.3	3512	1	X37085	FMF associated pro
18	282.6	9.2	1470	1	Q47355	Myotonic dystrophy
c 19	281.8	9.2	16891	1	X37084	MEFV gene sequence
20	281.6	9.2	8342	1	Q53478	MLL gene 8.3 kb Ba
21	281.6	9.2	8391	1	T16333	MLL gene 8.3 kb fr
22	281.4	9.1	24026	1	T18310	BRCA1, human breas
23	281.4	9.1	24026	1	T17455	Mutated BRCA1 geno
24	281.4	9.1	24026	1	T17512	Mutated BRCA1 geno
25	281.4	9.1	24026	1	T17513	Mutated BRCA1 geno
26	281.4	9.1	24026	1	T17514	Mutated BRCA1 geno
27	281.4	9.1	24026	1	T17515	Mutated BRCA1 geno
28	281.4	9.1	24026	1	T17516	Mutated BRCA1 geno
29	281.4	9.1	24026	1	T17517	Mutated BRCA1 geno
30	281.4	9.1	24026	1	T17518	Mutated BRCA1 geno
31	281.4	9.1	24026	1	T17519	Mutated BRCA1 geno
32	281.4	9.1	24029	1	T17520	Mutated BRCA1 geno
33	281.4	9.1	24026	1	T17521	Mutated BRCA1 geno
34	281.4	9.1	24026	1	T17522	Mutated BRCA1 geno
35	281.4	9.1	24026	1	T17523	Mutated BRCA1 geno
36	281.4	9.1	24026	1	T17524	Mutated BRCA1 geno
37	281.4	9.1	24031	1	T17525	Mutated BRCA1 geno
38	281.4	9.1	24026	1	T17526	Mutated BRCA1 geno
39	281.4	9.1	24026	1	T17527	Mutated BRCA1 geno
40	281.4	9.1	24026	1	T17528	Mutated BRCA1 geno
41	281.4	9.1	24026	1	T17529	Mutated BRCA1 geno
42	281.4	9.1	24026	1	T17530	BRCA1 genomic seq;
43	281.4	9.1	24026	1	T82612	BRCA1, human breas
c 44	280.2	9.1	4853	1	X91052	Nucleotide sequenc
45	280	9.1	9180	1	V57416	Tumour suppressor

ALIGNMENTS

RESULT 1
Q59527
ID Q59527 standard; cDNA; 384 BP.
AC Q59527;
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST00524.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; localities; chromosomes; chromosomal as.
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI; 93-272892/34.
PT Enriched oligonuclectides and corresp. sequences used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 189; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes. [illegible] of transcribed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST00524 has a "poor" coding probability as evaluated using the
CC coding-region prediction pr[illegible] TRM. See also [illegible].
SQ Sequence 384 BP; 72 A; 127 C; 96 G; 82 T;

Query Match 10.8%; Score 331.6; DB 1; Length 384;
Best Local Similarity 97.1%; Pred. No. 2.2e-65;
Matches 334; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1757 cgaggaaggcaagaagccccagggggtggcccgctgggtggtggtggcatgcctgcctgtgct 1816
Db 1 CGAGGAAGGCAAGAAGCCCACCCCCTGCCCCGCNTGGCGTCGGTGGCCTCCGTTCCTCTCT 60

Qy 1817 cgcagccccctatggtcagagctggatccaagattcaaaaccctttctttcctttcacccc 1876
Db 61 CGCAGCCCCTGTGGTCACAGCTCCATACAAGATTCTANCCCCTCTNTTCCCCCTNAACCC 120

Qy 1877 gctccaggttggagccacaga[illegible] 1936
Db 121 GCTCCAGGTTGGAGCCACAGACGCCACCCCCGCCCCCCCCTCCTCCNTCCTTTCCCTC 180

Qy 1937 tgccttccctccaagatgcgggcctcgacctgaatgaatc[illegible] 1996
Db 181 TGCCTTTCCCTCCAGAATGCGGCCTCGACCTGAATGAATCCCTTTCTCCCCCCTC 240

Qy 1997 gtcctgggctagctcctgacc[illegible] 2056
Db 241 GTCCTGGGGTAGCTCCTGACCTCTATTTTATTTTAAATTTTTTTTTTTTTTTA 300

Qy 2057 tttgacccgcccccttctcgctcataatgccaccgagtccttt 2100
Db 301 TTTGACCCGGCCCCTTCTCGCTCATAATGACCACGAGTTCCTT 344

RESULT 2

```
V29075/c
ID   V29075 standard; DNA; 3035 BP.
AC   V29075;
DT   11-SEP-1998  (first entry)
DE   Homo sapiens NGAL gene 5' region.
KW   5' region; NGAL gene; neu-related lipocalin; promoter; therapy;
KW   study; breast cancer; mammary gland; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   promoter        1102. .2910
FT                   /*tag= a
FT   misc_feature    2911
FT                   /*tag= b
FT                   /note= "putative transcription start site"
FT   TATA_signal     2892. .2888
FT                   /*tag= c
FT   CAAT_signal     2772. .2776
FT                   /*tag= d
PN   WO9815634-A1.
PD   16-APR-1998.
PF   05-MAY-1997; U07504.
PR   07-OCT-1996; US-726725.
PA   (WISC ) WISCONSIN ALUMNI RES FOUND
PI   Chen K, Gould MN;
DR   WPI, 98-240820/21.
PT   New isolated mammary gland specific promoters - can be used to
PT   promote gene expression in the mammary gland in a constant manner
PT   throughout the estrous cycle
PS   Disclosure; Fig 3; 48pp; English.
CC   The sequence is that of the 5' region of the human NGAL
CC   (also known as neutrophil lipocalin/HNL) gene. The sequence
CC   contains a mammary gland-specific promoter which can be used
CC   for targeting foreign gene expression to the mammary gland.
CC   It can be used in therapy and study of e.g. breast cancer genes.
SQ   Sequence  3035 BP;  624 A;  928 C;  809 G;  674 T;

  Query Match              9.8%;  Score 301.8;  DB 1;  Length 3035;
  Best Local Similarity   82.3%;  Pred. No. 1.6e-58;
  Matches  386;  Conservative    0;  Mismatches   72;  Indels   11;  Gaps    3;

Qy    2615 gctgggccaggcacgatggctcacgcctgtaatcccagactttgggaggccgaggcaggc 2674
           ||  |||| ||  || ||||||||  ||||||||||  ||||||||| ||||| ||
Db    1465 GCCAGGCCGGGTGCGGTGGCTCACATCTGTAATCCCAGCATTTGGGAGGCTGAGGCGGGT 1406

Qy    2675 ggatcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctct 2734
           |||||||  |||||  ||||||| |||  || |||||||||| ||||||| |||||||
Db    1405 GGATCAC--GAGGTCAGGAGTTCGAGACCAGTCTGGCCAACACGGTGAAACCCCATCTCT 1348

Qy    2735 actaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgg 2794
           ||||||||||||||||||  ||| |||||||| |||||||||||||||||||||||| ||
Db    1347 ACTAAAAATACAAAAATTAGCAGGGCATGGTGGCAGGCGCCTGTAATCCCAGCTACTCGG 1288

Qy    2795 gaggctgaggcatgagaatcacttaaacctgaagaggcagaggttacagtgagccaagatc 2854
           ||||||||||||  ||||   ||| |||||| ||||| |||||| ||||||||| |||||
Db    1287 GAGGCTGAGGCAGAAGAACAGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATC 1228

Qy    2855 gtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataaaaaa 2914
           |||  |||| |  |||||||   ||||||| | | |||||||  |||||||| ||||||
Db    1227 ACGCCTTTGCACTTTAGCCTGG--AACAGAGCGATATTCTGTCTTAAAAAAAAAAAAAAA 1170

Qy    2915 -------taatcaggcacgatggctcatgcctgtaatcccagcactctgggaggctgag 2967
                |  | |||| |||||||||  ||||||||  |||||||| |||||| |||||
Db    1169 GGTATCCTGGGCTGGGCGCAGTGGCTCTCGCCTGTAACTCCAGCACTTTGGGAGACTGAG 1110

Qy    2968 gtgggtggatcacctgaggtcaggagttcgagaccagcctggccaacatggcgaaaccccg 3027
           | |||||||| ||||||||||||| ||| ||||||||||  |||||||  |||||||||
Db    1109 GCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCCT 1050

Qy    3028 gtctctactaaaaatacgaaaaattagccgggcgtggtggtgcatgcctg 3076
           |||||||  |||||||||||| |||||  |||||||||| |||||||||
Db    1049 GTCTCTAGTAAAAATACAAACATTAGTTGGGCATGGTGGCGCATGCCTG 1001
```

```
RESULT   3
V99912
ID   V99912 standard; cDNA; 3099 BP.
AC   V99912;
DT   26-APR-1999  (first entry)
DE   Human sialoadhesin family 4 (SAF-1) related EST clone.
KW   SAF-4; sialoadhesin family; human; therapy; diagnosis; cancer;
KW   inflammation; autoimmune disease, allergy, asthma, inflammation;
KW   cerebellar degeneration; Alzheimer's disease; Parkinson's disease;
KW   multiple sclerosis, amyotrophic lateral sclerosis, head injury;
KW   septic shock; sepsis; stroke; osteoporosis; osteoarthritis;
KW   ischemia reperfusion injury; cardiovascular disease;
KW   kidney disease; liver disease; myocardial infarction; hypotension;
KW   hypertension; AIDS; myelodysplastic syndrome, aplastic anaemia;
KW   baldness; infection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             51. .1249
FT                   /*tag= a
FT                   /transl_except= (pos:1176. .1177, aa:Pro)
FT                   /note= "this codon has an apparent 1 nucleotide
FT                         deletion which alters the reading frame"
PN   WO9853840-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; U10791.
PR   27-MAY-1997; US-047572.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Erickson-Miller CL, Kikly KK;
DR   WPI; 99-080779/07.
DR   P-PSDB; W81024.
PT   New sialoadhesin family 4 polypetides and polynucleotides - useful
PT   to treat various diseases associated with SAF-4 expression
PS   Claim 13; Page 31-32; 48pp; English.
CC   This is the nucleotide sequence of an expressed sequence tag (EST)
CC   clone that encodes a polypeptide (see W81024) which shows close
CC   homology and/or structural similarity (e.g. a conservative amino
CC   acid difference) to new human sialoadhesin family 4 (SAF-4)
CC   polypeptide (see W81023).  SAF-4 polynucleotides and polypeptides,
CC   and methods for producing such polypeptides in transformed host
CC   cells using recombinant techniques, are disclosed.  SAF-4, its
CC   agonists and antagonists, and nucleic acid molecules that enhance
CC   or inhibit SAF-4 expression, may be used to treat patients in nee
CC   of enhancement or inhibition of SAF-4 expression or activity.
CC   Conditions that may benefit from such treatment include cancer,
CC   inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis,
CC   CNS inflammation, cerebellar degeneration, Alzheimer's disease,
CC   Parkinson's disease, multiple sclerosis, amyotrophic lateral
CC   sclerosis, head injury damage and other neurological disorders,
CC   septic shock, sepsis, stroke, osteoporosis, osteoarthritis,
CC   ischemia reperfusion injury, cardiovascular disease, kidney
CC   disease, liver disease, ischemic injury, myocardial infarction,
CC   hypotension, hypertension, AIDS, myelodysplastic syndromes and
CC   other haematologic abnormalities, aplastic anaemia, male baldness
CC   pattern and bacterial, protozoal, fungal and viral infections
CC   related to SAF-4 polypeptide activity.  Methods of identifying
CC   agonists, anatagonists/inhibitors are also provided, as well as
CC   diagnostic assays for detecting diseases associated with
CC   inappropriate SAF-4 activity or levels.
SQ   Sequence  3099 BP;  770 A;  907 C;  813 G;  607 T;

  Query Match              9.8%;  Score 300.6;  DB 1;  Length 3099;
  Best Local Similarity   79.8%;  Pred. No. 2.9e-58;
  Matches  395;  Conservative    3;  Mismatches   89;  Indels   11;  Gaps    3;

Qy    2585 ctccctgtctctggaagtatttaagaaaaggctgggcgcag--gcgtgatggctcacgcct 2642
           ||| || ||| |    | ||||| |||   |||| |  |  ||||  ||||||||||||
Db    2085 CTCTCTCTCTTTCTCTCTCTTTAAAAAACATCTGGCCAGGGAGTAGTAGTTTTTTTGTCT 2144

Qy    2643 gtaatcccag-actttgggaggccgaggcgggcggatcacctgaggtcaggagtttgagac 2701
           ||||||||| |||||||||||  |||| ||| |||| |||||||  |||||| |||||||
Db    2145 GTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCGAGAC 2204
```

```
Qy   2702 acagcctgggcaaacatggtgaaacctcatctctactaaaaatacaaaaattagccaggcg 2761
          ||||||||||||| |||| |||| | |||. ||||||||||||||||||||  |||
Db   2205 CCAGCCTGGGCAACTTGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCA 2264

Qy   2762 tggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgagaatcacttaaa 2821
          |||||||||| |||||||||| | |||||'||| |||||||| |||||||||| ||
Db   2265 TGGTGGCAGGGCGCCTGTAATCCCACCTACTTCGGAAGCTGAGGCAGGAGAATCACTTGAA 2324

Qy   2822 cctgaggaggcagaggttacagtgagccaagatcgtgccactgcattccagcctgggcgac 2881
          |||| ||| | |||||| ||||||||||||||   ||| |||| ||||||||||| ||
Db   2325 CCTGGGAGACGGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTGGGCAAC 2384

Qy   2882 agagcgagactctgtctcaaaaaaaaataaaaaataatcagggcacagtggctcatgcctg 2941
          | ||| ||||||  |||||||||||         ||||     | | |     || |||
Db   2385 AAAGCGAGACTCCATCTCAAAAAAAA--------AATCCTCCAAATGGGTTGGGTGTCTG 2436

Qy   2942 taatcccagcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcaagac 3001
          |||||||||||||| ||||||||| |||||' |||   | ||||  |||||||| ||||
Db   2437 TAATCCCAGCACTTTGGGAGGCTAAGGTGGGTGGATTGCTTGAGCCCAGGAGTTCGAGAC 2496

Qy   3002 cagcctggtgaacatggcgaaaccccgtctctaataaaaatacaaaaattagccgggcat 3061
          |||||||' |||||| |||||||| || ||'  ||||||||||||  |||| |||| |
Db   2497 CAGCCTGGGCAACATGGTGAAACCCCATCTCTACAAAAAATACAAAACATAGCTGGGCTT 2556

Qy   3062 ggtggtgcatgcctg 3076
          |||||||  ||||||
Db   2557 GGTGGTGTGTGCCTG 2571


RESULT   4
V99911
ID   V99911 standard; cDNA; 3099 BP.
AC   V99911;
DT   26-APR-1999  (first entry)
DE   Human sialoadhesin family 4 (SAF-1) cDNA.
KW   SAF-4; sialoadhesin family; human; therapy; diagnosis; cancer;
KW   inflammation; autoimmune disease; allergy; asthma; inflammation;
KW   cerebellar degeneration; Alzheimer's disease; Parkinson's disease;
KW   multiple sclerosis, amyotrophic lateral sclerosis; head injury;
KW   septic shock; sepsis; stroke; osteoporosis; ostecarthritis;
KW   ischemia reperfusion injury; cardiovascular disease;
KW   kidney disease; liver disease; myocardial infarction, hypotension;
KW   hypertension; AIDS; myelodysplastic syndrome; aplastic anaemia;
KW   baldness; infection; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             51..1970
FT                   /*tag= a
PN   WO9853840-A1.
PD   03-DEC-1998.
PF   27-MAY-1998; 98WO-US010791.
PP   27-MAY-1997; 97US-047572.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Erickson-Miller CL, Kikly KK;
DR   WPI; 99-080779/07.
DR   P-PSDB; W81023
PT   New sialoadhesin family 4 polypetides and polynucleotides - useful
PT   to treat various diseases associated with SAF-4 expression
PS   Claim 2; Page 30-31; 48pp; English.
CC   This cDNA clone codes for new human sialoadhesin family 4 (SAF-4)
CC   polypeptide (see W81023).  It can be obtained e.g. from a cDNA
CC   library derived from mRNA in cells of human primary dendritic
CC   cells, using expressed sequence tag analysis.  SAF-4
CC   polynucleotides and polypeptides, and methods for producing such
CC   polypeptides by recombinant techniques are disclosed.  SAF-4, its
CC   agonists and antagonists, and nucleic acid molecules that enhance
CC   or inhibit SAF-4 expression, may be used to treat patients in need
CC   of enhancement or inhibition of SAF-4 expression or activity.
CC   Conditions that may benefit from such treatment include cancer,
CC   inflammation, autoimmunity, allergy, asthma, rheumatoid arthritis,
CC   CNS inflammation, cerebellar degeneration, Alzheimer's disease,
CC   Parkinson's disease, multiple sclerosis, amyotrophic lateral
CC   sclerosis, head injury damage and other neurological disorders,
CC   septic shock, sepsis, stroke, osteoporosis, ostecarthritis,
CC   ischemia reperfusion injury, cardiovascular disease, kidney
CC   disease, liver disease, ischemic injury, myocardial infarction,
CC   hypotension, hypertension, AIDS, myelodysplastic syndroses and
CC   other haematologic abnormalities, aplastic anaemia, male baldness
CC   pattern and bacterial, protozoal, fungal and viral infections
CC   related to SAF-4 polypeptide activity.  Methods of identifying
CC   agonists  anatagonists/inhibitors are also provided as well as
CC   diagnostic assays for detecting diseases associated with
CC   inappropriate SAF-4 activity or levels.
SQ   Sequence   3099 BP;    769 A;    908 C;    816 G;    606 T;

  Query Match             9.8%;  Score 301.6;  DB 1;  Length 3099;
  Best Local Similarity   79.8%;  Pred. No. 2.9e-58;
  Matches  395;  Conservative    0;  Mismatches   89;  Indels   11;  Gaps    3;

Qy   2595 ctccctgtctctggaagtatttaagaaaaggctgggcgcagtggctcacgcctgtaatc 2642
          ||| || ||| |    | ||  |  ||  |  |  |  |  |  |  |  |  |  |
Db   2085 CTGTCTCTCTTTGTCTCTCTTTAAAAACATCTGGCTCACCGACGCCTGTAATCCC 2144

Qy   2643 gtaatcccag-actttgggaggccgaggcgggcggatcacctgaggtcaggagttcgagac 2701
          |||||||||| |||||||||  |  |  |  |  |  |  |  |  |  |  |  |||
Db   2145 GTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCGAGAC 2204

Qy   2702 acagcctggtgaacatggtgaa  |  |  |  |  |  |  |  |  |  |  |  |  | 2761
          | |||||||||| |||||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db   2205 CCAGCCTGGGCAACTTGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCA 2264

Qy   2762 tggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgagaatcacttaaa 2821
          |||||||||| |||||||||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db   2265 TGGTGGCAGGGCGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAA 2324

Qy   2822 cctgaggaggcagaggttacagtgagccaagatcgtgccactgcattccagcctgggcgac 2881
          |||| ||| | |||||| |||||||||||||  |  |  |  |  |  |  |  |  |
Db   2325 CCTGGGAGACGGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTGGGCAAC 2384

Qy   2882 agagcgagactctgtctcaaaaaaaaataaaaaataatcagggcacagtggctcatgcctg 2941
          | ||| ||||||  |||||||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db   2385 AAAGCGAGACTCCATCTCAAAAAAAA--------AATCCTCCAAATGGGTTGGGTGTCTG 2436

Qy   2942 taatcccagcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcaagac 3001
          |||||||||||||| |||||||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db   2437 TAATCCCAGCACTTTGGGAGGCTAAGGTGGGTGGATTGCTTGAGCCCAGGAGTTCGAGAC 2496

Qy   3002 cagcctggtgaacatggcgaaaccccgtctctaataaaaatacaaaaattagccgggcat 3061
          |||||||' |||||| |||||||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db   2497 CAGCCTGGGCAACATGGTGAAACCCCATCTCTACAAAAAATACAAAACATAGCTGGGCTT 2556

Qy   3062 ggtggtgcatgcctg 3076
          |||||||  ||||||
Db   2557 GGTGGTGTGTGCCTG 2571


RESULT   5
X23520
ID   X23520 standard; DNA; 45546 BP.
AC   X23520;
DT   23-JUN-1999  (first entry)
DE   Human kidney aminopeptidase P genomic DNA fragment 4.
KW   Aminopeptidase; human; AmP; gene therapy; treatment; AmP deficiency;
KW   prenatal diagnosis, angioedema, antihypertensive agent, atherosclerosis;
KW   arterial stenosis; industrial protein feed; malabsorption syndrome;
KW   proteinaceous waste degradation; additive; immunohistochemistry; ss.
OS   Homo sapiens.
PN   WO9911799-A2.
PD   11-MAR-1999.
PF   02-SEP-1998; 98WO-US018426.
PP   02-SEP-1997; 97US-057854.
PA   (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.
PI   Ryan JW, Sprinkle TJC, Venema RC;
DR   WPI; 99-205193/17.
```

```
PT   Nucleic acid encoding human aminopeptidase P
PS   Claim 13; Page 165-192; 201pp; English.
CC   This invention describes the isolation of a novel human aminopeptidase P
CC   (AmP). This protein is used to produce recombinant AmP and can be used
CC   for gene therapy for treating AmP-deficiency conditions. Its fragments
CC   are used as primers and probes to identify patients with homozygous and
CC   heterozygous AmP deficiency, including prenatal diagnosis (patients
CC   defective in AmP are at risk of developing angioedema if treated with
CC   angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC   in cases of excessive AmP expression. The product of the invention is
CC   also used to identify AmP-expressing sequences in other animals and to
CC   generate transgenic animals, and comparisons of genomic sequences are
CC   used to detect mutations. AmP inhibitors are potentially useful as
CC   antihypertensive agents and to prevent or treat arterial (re)stenosis
CC   or atherosclerosis. The structure of AmP is used to design synthetic
CC   substrates, e.g. for use in AmP assays. AmP, which hydrolyzes N-terminal
CC   imido bonds, can be used to degrade industrial protein feeds to free
CC   amino acids, to degrade proteinaceous wastes, as additives in enzyme
CC   formulations used to treat malabsorption syndrome and for studying its
CC   biological role. Antibodies against AmP are used in immunohistochemical
CC   methods to study AmP distribution.
SQ   Sequence   45546 BP;    12027 A;    11359 C;    11380 G;    10780 T;

  Query Match              9.6%;  Score 205.8;  DB 1;  Length 45546;
  Best Local Similarity   81.6%;  Pred. No. 6.3e-57;
  Matches  394;  Conservative    0;  Mismatches   77;  Indels   12;  Gaps    4;

Qy    2605 ttaagaaaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggagg 2663
           |||| ||| | | || ||| | ||||| | ||||||||||||||| ||||||| |||
Db   25149 TTAAAGAAAATACAAGACCGGGGGTGGTGGCTTATGCCTGTAATCCCAGCACTTTGGAAGG 25208

Qy    2664 ccgaggcgggcggatcacctgaggtcaggagttcgagaccagcctggccaacatggtgaa 2723
           ||||||| || |||||||||||||| |||||| |||| ||||||| |||||| || |||
Db   25209 CCGAGGCTGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACACGGAGAA 25268

Qy    2724 accccatctctactaaaaatacaaaaattagccgggcgtggtggcaggtgcctgtaatcc 2783
           ||| ||| |||||||||||||| |||||| ||  || |||||||    ||||||||||||
Db   25269 ACCCCATATCTACTAAAAATAC-AAAATTAG-CGGGTGTGGTGGTGCATGCCTGTAATCC 25326

Qy    2784 agctactcgggaggctgaggcatgagaattgcttgaacctgggaggcagaggttacagt 2843
           |||||||  | |||||||| |||| || ||  |||||| |||| |||||| | ||
Db   25327 CAGCTACTCAGGGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTGCAGT 25386

Qy    2844 gagccaagatcgtgccactgcattccagcctgggcgac-agagcaagactctgtctcaaa 2902
           |||||||'|||| |||| || | |||| ||'|||| || ||||||| ||||||||||||
Db   25387 GAGCCAAGATCATGCCATTGCACTCCAACCTGGGCAACAAGAGCAAAACTCTGTCTCAAA 25446

Qy    2903 aaaaataaaaaataa---- ----tcaggaggctgagtggctcatgcctgtaatcccagcac 2953
           ||||| |||||| |          |  ||||  ||||||||  ||||||||||||||  ||
Db   25447 AAAAAAAAAAAAATATAAATCAGGTCGATGGTGGCTCACACCTGTAATCCCAATAC 25507

Qy    2954 tctgggaggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaa 3013
           | |||||||| || ||  | ||||||| |||  ||||||||||||||||||||||  ||
Db   25508 TTTGGGAGGCAGAGGTCGGCGATCACTTGAGTTCAGGAGTTCAAGACCAGCCTGGGCAA 25567

Qy    3014 catggtgaaaccccgtctctactaaaaatacaaaaattagccgggcgtggtggtgcatgc 3073
           |||| ||| ||| | |||| |||||| ||  |||||||| ||| |||| |||  |||
Db   25568 CATGGTGAAGCCCATTTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTAGTGTGTGC 25627

Qy    3074 ctg 3076
           |||
Db   25628 CTG 25630

RESULT   6
T18551/c
ID   T18551 standard; DNA; 53577 BP.
AC   T18551;
DT   06-MAY-1997  (first entry)
DE   Human polycystic kidney disease normal PKD1 gene.
KW   Adult onset polycystic kidney disease; APKD; autosomal dominant;
KW   mutant; transversion; transition; deletion; insertion; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_feature    4379. .5272
FT                   /*tag= a
FT                   /note= "specifically claimed region of intronless
FT                   cDNA identified by exon trapping"
FT   old_sequence    replace(50652. .50653, cg)
FT                   /*tag= b
FT                   /note= "changes Val codon to Leu codon"
FT   old_sequence    replace(50796. .50797, cg)
FT                   /*tag= c
FT                   /note= "replaces Val codon by Leu codon"
FT   old_sequence    insert(51827, 51828, cc)
FT                   /*tag= d
FT                   /note= "insertion, results in frameshift"
PN   WO9612033-A1.
PD   25-APR-1996.
PF   11-OCT-1995; U13357.
PR   12-OCT-1994; US-323443.
PR   31-JAN-1995; US-381520.
PA   (IGIG-) IG LAB INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Burn TC, Connors TD, Dackowski W, Germino G, Klinger KW;
PI   Landes GM, Qian F;
DR   WPI; 96-222017/22.
PT   Isolated human polycystic kidney disease gene and its mutants -
PT   useful for treatment of polycystic kidney disease and screening for
PT   carriers
PS   Claim 1; Fig 1; 65pp, English.
CC   The present sequence is that of the normal human PKD1 gene from
CC   chromosome 16. Mutations in this gene (e.g. transitions,
CC   transversions, deletions and/or insertions) are associated with
CC   adult-onset polycystic kidney disease (APKD). The PKD1 locus is
CC   GC-rich (62.4%). Comparison of this sequence with a previously
CC   reported partial cDNA sequence revealed differences at three
CC   locations (see features table). The most significant difference is
CC   the presence of two additional cytosine residues on the plus-strand
CC   at position 4566 of the previously reported sequence. The insertion
CC   results in a frame-shift in the predicted protein sequence,
CC   leading to replacement of 92 C-terminal amino acids with a novel
CC   12 amino acid C-terminus. The PKD1 gene contains 33 Alu repeats.
CC   There is a region consisting of 17 tandem copies of a perfect 27 bp
CC   repeat and two large CT-rich regions.
SQ   Sequence   53577 BP;    8495 A;    17681 C;    15785 G;    11616 T;

  Query Match              9.5%;  Score 203.2;  DB 1;  Length 53577;
  Best Local Similarity   82.9%;  Pred. No. 2.5e-56;
  Matches  383;  Conservative    0;  Mismatches   73;  Indels    5;  Gaps    4;

Qy    2610 aaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccga 2668
           | || ||  |||||||||| |||||||| |||||||||||||| ||||||  | | |
Db    1295 AGAAAGCCAGGCCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTCGGAGGCCAA- 1353

Qy    2669 gcaggcggatcacctgaggtgaggagttcgagaccagcctgaccaacatgttgaaaccc- 2728
           |||||||| |||| |||||  | |||| |||| || |||||||| ||| || | | |
Db    1235 GCAGGCGGGTCACATGAGGCCAAGACTTCGAGACCAACCTGGCCAACGTGGTGAAACCCC 1176

Qy    2729 atctctactaaaaatacaaaaattagccgggcgtggtggcaggtgcctgtaatcccagct 2788
           |||||||||'|||||||||||||||||||||| |||||  ||||||||||||||||||||
Db    1175 GTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGGGTGCCTGTAATCCCAGCT 1116

Qy    2789 acttgggaggctgaggcatgagaattgcttgaacctgggaggcggaggttacagtgagcc 2848
           ||| |||||.|||||| ||||||||| |||| | |||| |||| |||||| | ||||
Db    1115 ACTCGGGAGGCTGAGGCAGGAGAATCACTGGAACCCGGGAGGCGGAGGTCGCAGTGAGCC 1056

Qy    2849 aagatcgtgccactgcattccagcctgggcgacagagcgagactccgtctcaaaaaaat 2908
           || |   |.|||||| |||||| | |||| |  |||| |||||||||||||||||||
Db    1055 GAGGTTACCCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAGA-- 997

Qy    2909 aaaaaataatcagggcacagtggctcatgcctgtaatcccagcactttgggaggccgag 2968
           |||||| |  | |||||||||||||||||||||||||||||||||||| |||| | |||
Db     997 AAAAAAAAGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAG 938
```

```
Qy   2969 tgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccg 3028
          ||||||||||  |||||||  ||| |  || ||| |||||  |  |||| |||||||||
Db    937 CAGGTGGATCAC--GAGGTCAAGAGATTCAG-CCATCCTGGCCAGTATGGTGAAACCCCG 881

Qy   3029 tctctaataaaaatacaaaaattagccgggcatggtggtgca 3070
          | |||' |||||||||||||||| || ||||  |||||  ||
Db    880 TGTCTACTAAAAATACAAAAATTGGCTGGGCATGGTGGCACA 839


RESULT   7
T94108/c
ID   T94108 standard; DNA; 53577 BP.
AC   T94108;
DT   01-JUN-1998  (first entry)
DE   Human PKD1 locus between chromosomal markers ATPL (ATP6C) and D16S84.
KW   Human; polycystic kidney disease 1; PKD1; treatment;
KW   autosomal dominant polycystic kidney disease; APKD; ss.
OS   Homo sapiens.
PN   WO9744457-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; 97WO-US08799.
PR   03-JUN-1996; US-658136.
PR   24-MAY-1996; US-655360.
PA   (GENZ ) GENZYME CORP.
PI   Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI   Qian F;
DR   WPI; 98-018511/02.
PT   Human polycystic kidney disease gene, PKD1 - useful to treat and
PT   diagnose human autosomal or adult onset polycystic kidney disease
PS   Example 5; Pages 60-89; 257pp; English.
CC   The present sequence is the human polycystic kidney disease 1
CC   (PKD1) locus between chromosomal markers ATPL (ATP6C) and D16S84.
CC   The PKD1 gene or polypeptide may be used to treat autosomal
CC   dominant polycystic kidney disease (APKD), and identify carriers
CC   of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC   (Ab) that distinguish between normal and mutant PKD1 sequences can
CC   also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC   perform subcellular and histochemical localisation studies, and to
CC   block the function of PKD1. Ab are also useful in rational drug
CC   design studies to identify and test inhibitors of PKD1. Sense and
CC   antisense sequences derived from the PKD1 gene may used for
CC   detection and therapy.
SQ   Sequence   53577 BP;    8495 A;    17684 C;    15782 G;    11616 T;


  Query Match             9.5%;  Score 293.2;  DB 1;  Length 53577;
  Best Local Similarity   82.9%;  Pred. No. 2.5e-56;
  Matches  383;  Conservative    0;  Mismatches   73;  Indels    6;  Gaps    4;

Qy   2610 aaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgag 2668
          | || ||  ||||||||||  |||||||| | ||||||||||| ||||||||||||| ||
Db   1295 AGAAAGCCAGGCCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAG 1236

Qy   2669 gcaggcggatcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctc 2728
          |||||||| |||| |||||  | ||||| |||| || ||||||||||  |||  ||||| |
Db   1235 GCAGGCGGGTCACATGAGGCCAAGAGTTCGAGACCAACCTGGCCAACGTGGCAAAACCCC 1176

Qy   2729 atctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagct 2788
          ||||||||||||||||||||||||||||||| ||||||  ||||||||||||||||||||
Db   1175 GTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGGGTGCCTGTAATCCCAGCT 1116

Qy   2789 acttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagcc 2848
          ||| |||||||||||||| ||||||||||  | || | ||||| |||||  |||||| |
Db   1115 ACTCGGGAGGCTGAGGCAGGAGAATCACTGGAACCCGGGAGGCGGAGGTCGCAGTGACCT 1056

Qy   2849 aagatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaat 2908
          || |   |||||||| ||||||||||| | |||||  ||||||  |||||||||| | |
Db   1055 GAGGTTACCCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAGA-- 997

Qy   2909 aaaaaataatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgagg 2968
          |||||| |  | |||||||||||||||| | ||||||||||||| |||||||| ||||
Db    997 AAAAAAAAGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCGGAGG 938
```

```
Qy   2969 tgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccg 3028
          ||||||||||  |||||||  ||| |  || ||| |||||  |  |||| |||||||||
Db    937 CAGGTGGATCAC--GAGGTCAAGAGATTCAG-CCATCCTGGCCAGTATGGTGAAACCCCG 881

Qy   3029 tctctaataaaaatacaaaaattagccgggcatggtggtgca 3070
          | |||' |||||||||||||||| || ||||  |||||  ||
Db    880 TGTCTACTAAAAATACAAAAATTGGCTGGGCATGGTGGCACA 839


RESULT   8
I94101/c
ID   I94101 standard; DNA; 53526 BP.
AC   I94101;
DT   01-JUN-1998  (first entry)
DE   Human PKD1 gene.
KW   Human; polycystic kidney disease 1; PKD1; treatment;
KW   autosomal dominant polycystic kidney disease; APKD; ss.
OS   Homo sapiens.
PN   WO9744457-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; 97WO-US08799.
PR   03-JUN-1996; US-658136.
PR   24-MAY-1996; US-655360.
PA   (GENZ ) GENZYME CORP.
PI   Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI   Qian F;
DR   WPI; 98-018511/02.
PT   Human polycystic kidney disease gene, PKD1 - useful to treat and
PT   diagnose human autosomal or adult onset polycystic kidney disease
PS   Claim 2; Pages 90-118; 257pp; English.
CC   The present sequence is the human polycystic kidney disease 1
CC   (PKD1) gene. The PKD1 gene or polypeptide may be used to treat
CC   autosomal dominant polycystic kidney disease (APKD), and identify
CC   carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
CC   Antibodies (Ab) that distinguish between normal and mutant PKD1
CC   sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
CC   also be used to perform subcellular and histochemical localisation
CC   studies, and to block the function of PKD1. Ab are also useful in
CC   rational drug design studies to identify and test inhibitors of
CC   PKD1. Sense and antisense sequences derived from the PKD1 gene may
CC   used for detection and therapy.
SQ   Sequence   53526 BP;    8486 A;    17665 C;    15769 G;    11606 T;


  Query Match             9.5%;  Score 293.2;  DB 1;  Length 53526;
  Best Local Similarity   82.9%;  Pred. No. 2.5e-56;
  Matches  383;  Conservative    0;  Mismatches   73;  Indels    6;  Gaps    4;

Qy   2610 aaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgag 2668
          | || ||  ||||||||||  |||||||| | ||||||||||| ||||||||||||| ||
Db   1295 AGAAAGCCAGGCCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAG 1236

Qy   2669 gcaggcggatcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctc 2728
          |||||||| |||| |||||  | ||||| |||| || ||||||||||  |||  ||||| |
Db   1235 GCAGGCGGGTCACATGAGGCCAAGAGTTCGAGACCAACCTGGCCAACGTGGCAAAACCCC 1176

Qy   2729 atctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagct 2788
          ||||||||||||||||||||||||||||||| ||||||  ||||||||||||||||||||
Db   1175 GTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGGGTGCCTGTAATCCCAGCT 1116

Qy   2789 acttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagcc 2848
          ||| |||||||||||||| ||||||||||  | || | ||||| |||||  |||||| |
Db   1115 ACTCGGGAGGCTGAGGCAGGAGAATCACTGGAACCCGGGAGGCGGAGGTCGCAGTGACCT 1056

Qy   2849 aagatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaat 2908
          || |   |||||||| ||||||||||| | |||||  ||||||  |||||||||| | |
Db   1055 GAGGTTACCCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAGA-- 997

Qy   2909 aaaaaataatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgagg 2968
          |||||| |  | |||||||||||||||| | ||||||||||||| |||||||| ||||
Db    997 AAAAAAAAGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCGGAGG 938
```

```
Qy   2969 tgagtgaatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccg 3028
          ||||||||||  |||||||  ||| |  ||  || |||||  |  |||| |||||||||
Db    937 CAGGTGGATCAC--GAGGTCAAGAGATTGAG-TCATCCTGGCCAGTATGGTGAAACCCCG 881

Qy   3029 tctctaataaaaatacaaaaattagccgggcatggtggtgca 3070
          | |||| ||||||||||||||||| || |||| ||||||  ||
Db    880 TGTCTACTAAAAATACAAAAATTGGCTGGGCATGGTGGCACA 839


RESULT  9
V27941/c
ID   V27941 standard; cDNA; 14796 BP.
AC   V27941;
DT   25-SEP-1998  (first entry)
DE   Survivin gene.
KW   survivin; apoptosis, cellular apoptosis, transplantation, ss,
KW   motor neuron degenerative disease; HIV infection; immunosuppression;
KW   gastrointestinal perturbations, cardiovascular disorder.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             2811. .12044
FT                   /*tag= a
FT                   /product= Survivin
FT   exon            2811. .2921
FT                   /*tag= b
FT                   /number= 1
FT   intron          2922. .3173
FT                   /*tag= c
FT                   /number= 1
FT   exon            3174. .3283
FT                   /*tag= d
FT                   /number= 2
FT   intron          3284. .5157
FT                   /*tag= e
FT                   /number= 2
FT   exon            5158. .5275
FT                   /*tag= f
FT                   /number= 3
FT   intron          5276. .11954
FT                   /*tag= g
FT                   /number= 3
FT   exon            11955. .12044
FT                   /*tag= h
FT                   /number= 4
PN   WO9822589-A2.
PD   28-MAY-1998.
PF   20-NOV-1997; U21880.
PR   20-NOV-1997; US-975080.
PR   20-NOV-1996; US-031435.
PA   (UYYA ) UNIV YALE.
PI   Altieri DC;
DR   WPI; 98 312475/27.
DR   P-PSDB; W614359.
PT   Modulating apoptosis by controlling the Survivin gene - useful for
PT   treating transplant rejection, degenerative disorders and tumours
PS   Disclosure; Fig 10; 108pp; English.
CC   The survivin gene can be used to control apoptosis through modification
CC   of the gene.  Survivin peptides can be used to inhibit cellular
CC   apoptosis, e.g. for enhancing the viability of organs and tissues prior
CC   to their transplantation, for preserving the growth of cells in culture
CC   or for treating conditions involving abnormal apoptosis,
CC   e.g. degenerative diseases such as motor neuron degenerative diseases,
CC   HIV infection, dermatological effects of ageing, disorders and diseases
CC   such as immunosuppression, gastrointestinal perturbations, cardiovascular
CC   disorders, apoptosis related to reperfusion damage, rejection of tissue
CC   transplantation and Alzheimer's disease.  Agents which block Survivin
CC   activity can be used to treat e.g. tumours
SQ   Sequence  14796 BP;  3482 A;  3438 C;  3888 G;  3938 T;


  Query Match             3.5%;  Score 293;  DB 1;  Length 14796;
  Best Local Similarity   81.7%;  Pred. No. 2.1e-56;
  Matches  375;  Conservative    0;  Mismatches   80;  Indels    4;  Gaps    3;

Qy   2619 ggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgag-ggggtggga 2677
          |||||||   | ||||||| || |||| || |||| |||||| ||| . ||
Db   5969 GGCCAGGTGTGGTGGCTCACACCTATAATCTCAGCACTTTGGGAGGCTGAGGTGGGTGGA 5910

Qy   2678 tcacctgaggtgaggagtttgaaaaccagcctggccaacatggtgaaaccctgtctctact 2737
          |||||||||||| ||||||| | | ||||||||||||||||||||||||||| ||||||
Db   5909 TCACCTGAGGTCAGGAGTTCAAAACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACT 5850

Qy   2738 aaaaatacaaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggag 2797
          ||||||||||||||| |  || |||||||  |||||||||| | ||||||||||
Db   5849 AAAAATACAAAAATTACCTGGGTGTGGTGGCACATGCCTGTAATCCCGGCTACTTGGGAG 5790

Qy   2798 gctgaggcatgagaatcacttaacctgaggaggcagaggttacagtgagccgagatcgtg 2857
          ||||| ||| | |||||||| | || |||| |||||| |||||| || | |.
Db   5789 GCTGAAGCAGGCAAATCACTTGAGCCTAGGAGGCGGAGGTTGCAGTGAGCCGAGATCAGC 5730

Qy   2858 ccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaataaaaaataa 2917
          ||||  | ||||||||||| |||||  | ||| ||||||| ||  || ||||| ||
Db   5729 CCACGCCTACTCCAGCCTGGGTAACAGAGTGAAACTGTGTCTCAGAACAAA-AAAAAAAAA 5671

Qy   2918 tcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtggat 2977
          | |||  |||| ||| |||||||||||||||| |||||||  ||| ||| |
Db   5670 GCCAGGCGTGGTGGTTCACGCCTGTAATCCCAGCACTTTGGGAGGCTAAGTTGGGTGGAT 5611

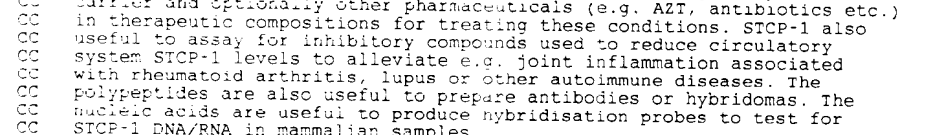
Qy   2978 cacctgaggtcaggagttcaagaccagcctggccaacatggtgaaaccccgtctctacta 3037
          |||  |||||||||||||||||||||||||  ||  ||| ||||||| ||||| ||
Db   5610 CAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTA 5553

Qy   3038 aaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          ||||||||||||||||| |||  |||||  | | ||||
Db   5552 AAAATACAAAAATTAGCTGGGTGTGGTGGCACTTCCCTG 5514


RESULT  10
V38933/c
ID   V38933 standard; DNA; 7146 BP.
AC   V38933;
DT   24-SEP-1998  (first entry)
DE   Nucleic acid sequence of genomic DNA encoding human STCP-1.
KW   Human; STCP-1; chemokine activity; T-cells; treatment; HIV infection;
KW   inhibitory compound, assay, reduce, circulatory system STCP 1 level;
KW   joint inflammation; rheumatoid arthritis; lupus; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1437. .6204
FT                   /*tag= a
FT                   /product= STCP_1
FT                   /note= "contains introns and exons"
FT   exon            1437. .1509
FT                   /*tag= b
FT                   /number= 1
FT   exon            1510. .3056
FT                   /*tag= c
FT                   /number= 1
FT   exon            3057. .3180
FT                   /*tag= d
FT                   /number= 2
FT   intron          3181. .6119
FT                   /*tag= e
FT                   /number= 2
FT   exon            6120. .6201
FT                   /*tag= f
FT                   /number= 3
PN   WO9824907-A1.
PD   11-JUN-1998.
PF   26-NOV-1997; U21552.
PR   03-DEC-1996; US-760127.
PA   (AMGE-) AMGEN INC.
PI   Andrew DP, Chang M;
DR   WPI; 98-333326/29.
DR   P-PSDB; W62783.
PT   Human STCP-1 polypeptides with chemokine activity - useful e.g. to
```

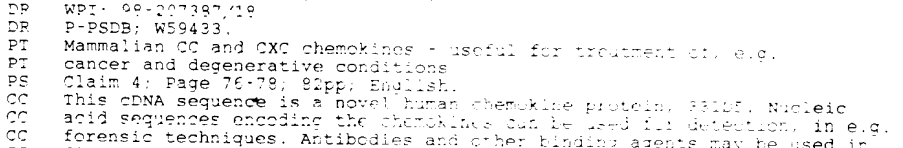
PT treat HIV infection or other viral or bacterial pathogens infecting
PT T-cells, macrophages or other immune system cells
PS Claim 1; Fig 2A-F; 96pp; English.
CC The present sequence encodes human STCP-1. STCP-1 polypeptides
CC demonstrate chemokine activity for T-cells. The polypeptides are useful
CC prophylactically or therapeutically to treat HIV infection and other
CC conditions associated with viral/bacterial pathogens infecting T-cells,
CC macrophages or other immune system cells. They can be included
CC (optionally chemically modified) with a pharmaceutically acceptable
CC carrier and optionally other pharmaceuticals (e.g. AZT, antibiotics etc.)
CC in therapeutic compositions for treating these conditions. STCP-1 also
CC useful to assay for inhibitory compounds used to reduce circulatory
CC system STCP-1 levels to alleviate e.g. joint inflammation associated
CC with rheumatoid arthritis, lupus or other autoimmune diseases. The
CC polypeptides are also useful to prepare antibodies or hybridomas. The
CC nucleic acids are useful to produce hybridisation probes to test for
CC STCP-1 DNA/RNA in mammalian samples.
SQ Sequence 7146 BP; 1664 A; 1814 C; 1883 G; 1785 T;

Query Match 9.5%; Score 293; DB 1; Length 7146;
Best Local Similarity 79.3%; Pred. No. 1.8e-56;
Matches 380; Conservative 0; Mismatches 80; Indels 19; Gaps 2;



RESULT 11
V34997/c
ID V34997 standard; cDNA; 1930 BP.
AC V34997;
DT 27-AUG-1998 (first entry)
DE Human chemokine protein 331D5 cDNA.
KW Chemokine; human; detection; forensic; diagnostic; treatment; cancer;
KW degenerative condition, abnormal proliferation; regeneration;
KW degeneration; atrophy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 48. .329
FT /*tag= a
FT sig_peptide 48. .119
FT /*tag= b
FT mat_peptide 120. .326
FT /*tag= c
FT /product= "chemokine protein 331D5"
PN WO9811226-A2.
PD 19-MAR-1998.
PF 09-SEP-1997; U15315.
PR 10-SEP-1996; US-025724.
PA (SCHE) SCHERING CORP.
PT Gorman DM, Hedrick JA, Zlotnik A;
DR WPI; 98-207387/19.
DR P-PSDB; W59433.
PT Mammalian CC and CXC chemokines - useful for treatment of, e.g.
PT cancer and degenerative conditions
PS Claim 4; Page 76-78; 82pp; English.
CC This cDNA sequence is a novel human chemokine protein, 331D5. Nucleic
CC acid sequences encoding the chemokines can be used for detection, in e.g.
CC forensic techniques. Antibodies and other binding agents may be used in
CC diagnostics. The chemokines themselves are useful for treatment of, e.g.
CC cancer or degenerative conditions. Abnormal proliferation, regeneration,
CC degeneration or atrophy may be treated by the inventive compositions.
SQ Sequence 2930 BP; 596 A; 867 C; 677 G; 790 T;

Query Match 9.3%; Score 286.6; DB 1; Length 2930;
Best Local Similarity 76.1%; Pred. No. 3.9e-55;
Matches 404; Conservative 0; Mismatches 99; Indels 28; Gaps 3;



RESULT 12
T76529/c
ID T76529 standard; DNA; 2923 BP.
AC T76529;
DT 11-SEP-1997 (first entry)
DE DNA encoding macrophage derived chemokine for treating inflammation.
KW MDC; macrophage derived chemokine; C-C; Cys-Cys; Crohn's disease;
KW rheumatoid arthritis; chemotaxis; fibroblast proliferation;
KW wound healing; angiogenesis; atherosclerosis; inflammation; ss.
OS Homo sapiens.

```
FH   Key             Location/Qualifiers
FT   cds             20. .301
FT                   /*tag= a
FT   signal_peptide  20. .91
FT                   /*tag= b
FT   mat_peptide     92. .298
FT                   /*tag= c
PN   WO9640923-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U10114.
PR   07-JUN-1995; US-479620.
PR   16-NOV-1995; US-558658.
PA   (ICOS-) ICOS CORP.
PI   Godiska R, Gray PW;
DR   WPI; 97-052324/05.
DR   P-PSDB; W20058.
PT   Macrophage derived chemokine (MDC) and analogues - used in the
PT   treatment of inflammatory diseases, MDC antibodies used to treat
PT   Crohn's disease, rheumatoid arthritis, etc.
PS   Claim 3; Page 71-72; 106pp; English.
CC   A new macrophage derived chemokine, MDC, a member of the C-C
CC   (Cys-Cys) subfamily of cytokines has been isolated. MDC and it's
CC   analogues may be used in the treatment of inflammatory diseases
CC   especially diseases characterised by monocyte chemotaxis towards a
CC   site of inflammation. MDC and it's analogues also induce fibroblast
CC   proliferation having a positive effect in wound healing and
CC   angiogenesis. They may prove to be clinically important in the
CC   treatment of tumours, by directly or indirectly inhibiting tumour
CC   formation. Antibodies directed against MDC and its analogues may be
CC   used in the treatment of Crohn's disease, rheumatoid arthritis and
CC   atherosclerosis. Probes and/or primers for the identification of MDC
CC   encoding sequences can be derived from MDC encoding sequences.
SQ   Sequence   2923 BP;    605 A;    861 C;    669 G;    788 T;
```

```
  Query Match             9.3%;  Score 285.6;  DB 1;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No. 6.4e-55;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy    2573 agaagcactgagctccctgtctctggaagtatttaagaaaaggctgggccaggcacgatg 2632
           |||| | | | | | ||| | | || ||| || ||| ||  ||| ||   |||
Db    1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy    2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
           |||||||||||||||| ||| |||||||||  | ||| ||| ||||||| |||||| ||
Db    1805 GCTCACGCCTGTAATCTCAGCACTTTGGGAGCACAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy    2692 gagtttgagaaacagcctggccaacatggtgaaacctcatctctactaaaaatacaaaaat 2751
           |||||  ||| |||||||||||| |||||||  ||||  ||||||||||||||||||||
Db    1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-GTCTCTACTAAAAATACAAAAAT 1687

Qy    2752 tagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgaga 2811
           |||||| ||| ||||||    ||||||||||  |||||||||||||||||  ||    ||||
Db    1686 TAGCCAGCCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGGGGTGGGAGA 1627

Qy    2812 atcacttaaacctgagaggcagaggttacagtgagccaagatcgtgccactgcattccag 2871
           |||||| ||||  ||||  |||||  || ||||| |||||  ||||||||| |||||
Db    1626 ATCACTTGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy    2872 cctgggcgacagagcaagactctgtctcaaaaaaaataaaaaat·················· 2916
           |||||||||||||  |||| ||| |  ||| ||| ||||||
Db    1566 CCTGGGCGACAGAGTGAGACCCTGCCATAAAAAAAAAAAAAAAAGTTGGGGGGGGCAGTGG 1507

Qy    2916 ·· ------- aatcagggcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                |  ||||| |||||||  |||||||||||||||| ||||||||
Db    1506 GGAGAGGGAAGAGGCCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447

Qy    2965 gaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaac 3024
           |||| |.|| |||| |||||||| ||||||||  ||||||||||  |||||||  ||| |
Db    1446 GAGGTGGGCGGATTACCTGAGGACAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAGC 1387

Qy    3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | |||||||| ||||||||||||||||||  ||| |||| |      ||||
```

```
Db    1386 CTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTAGCAGGAACCTG 1335


RESULT  13
T99233/c
ID   T99233 standard; cDNA; 2923 BP.
AC   T99233;
DT   01-APR-1998  (first entry)
DE   Coding sequence for macrophage-derived chemokine.
KW   Macrophage-derived chemokine; MDC; antibody; binding modulator; therapy;
KW   arthritis, inflammatory disorder, cancer, Crohn's disease;
KW   atherosclerosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             20. .301
FT                   /*tag= a
PN   US5688927-A.
PD   18-NOV-1997.
PF   07-JUN-1995; 480449
PR   07-JUN-1995; US-480449.
PA   (ICOS-) ICOS CORP
PI   Godiska R, Gray PW;
DP   WPI; 98-008038/01
DP   P-PSDB; W40811.
PT   Antibodies specific for macrophage-derived chemokine - useful for
PT   purifying or detecting the chemokine or modulating its activity
PS   Disclosure; Column 21-24; 22pp; English
CC   This sequence represents the coding sequence for a macrophage-derived
CC   chemokine (MDC). The encoded protein is used to produce the antibodies of
CC   the invention. The antibodies are useful for purifying MDC polypeptides,
CC   for detecting endogenous MDC in a host, and for modulating binding of MDC
CC   to its receptors. This sequence can be used for identifying and isolating
CC   non-human MDC homologues. The encoded protein is potentially useful for
CC   treating inflammatory disorders, cancer, etc. Antagonists of MDC can be
CC   used for treating Crohn's disease, arthritis, atherosclerosis etc.
SQ   Sequence   2923 BP;    605 A;    861 C;    669 G;    788 T;
```

```
  Query Match             9.3%;  Score 285.6;  DB 1;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No. 6.4e-55;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy    2573 agaagcactgagctccctgtctctggaagtatttaagaaaaggctgggccaggcacgatg 2632
           |||| | | | | | ||| | | || ||| || ||| ||  ||| ||   |||
Db    1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy    2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
           |||||||||||||||| ||| |||||||||  | ||| ||| ||||||| |||||| ||
Db    1805 GCTCACGCCTGTAATCTCAGCACTTTGGGAGCACAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy    2692 gagtttgagaaacagcctggccaacatggtgaaacctcatctctactaaaaatacaaaaat 2751
           |||||  ||| |||||||||||| |||||||  ||||  ||||||||||||||||||||
Db    1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-GTCTCTACTAAAAATACAAAAAT 1687

Qy    2752 tagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgaga 2811
           |||||| ||| ||||||    ||||||||||  |||||||||||||||||  ||    ||||
Db    1686 TAGCCAGCCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGGGGTGGGAGA 1627

Qy    2812 atcacttaaacctgagaggcagaggttacagtgagccaagatcgtgccactgcattccag 2871
           |||||| ||||  ||||  |||||  || ||||| |||||  ||||||||| |||||
Db    1626 ATCACTTGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy    2872 cctgggcgacagagcaagactctgtctcaaaaaaaataaaaaat·················· 2916
           |||||||||||||  |||| ||| |  ||| ||| ||||||
Db    1566 CCTGGGCGACAGAGTGAGACCCTGCCATAAAAAAAAAAAAAAAAGTTGGGGGGGGCAGTGG 1507

Qy    2916 ·· ------- aatcagggcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                |  ||||| |||||||  |||||||||||||||| ||||||||
Db    1506 GGAGAGGGAAGAGGCCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447

Qy    2965 gaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaac 3024
           |||| |.|| |||| |||||||| ||||||||  ||||||||||  |||||||  ||| |
Db    1446 GAGGTGGGCGGATTACCTGAGGACAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAGC 1387
```

```
Qy    3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | ||||||| |||||||||||||||||||  ||| |||| |     ||||
Db    1386 CTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTAGCAGGAACCTG 1335


RESULT  14
X32817/c
ID   X32817 standard; DNA; 2923 BP.
AC   X32817;
DT   25-JUN-1999  (first entry)
DE   Macrophage derived chemokine (MDC) encoding DNA.
KW   Macrophage derived chemokine; MDC; lentivirus infection; human; HIV;
KW   human immunodeficiency virus; feline immunodeficiency virus;
KW   bovine immunodeficiency virus; ss.
OS   Homo sapiens.
PN   WO9914237-A1.
PD   25-MAR-1999.
PF   16-SEP-1998; U19450.
PR   16-SEP-1997; US-931764.
PA   (ALKU ) AKZO NOBEL NV.
PI   Devico AL, Gallo RC, Garzino-Demo A, Markham PD,
PI   Pal R;
DR   WPI; 99-244024/20.
DR   P-PSDB; Y06829
PT   Treatment or prevention of lentivirus, particularly HIV infection
PS   Examples; Page 97-98; 103pp; English.
CC   This DNA encodes a human macrophage derived chemokine (MDC). The
CC   invention provides a novel method of treating or preventing lentivirus
CC   (LV) infection or replication in a human subject, that comprises
CC   administering to the subject a composition comprising MDC or a derivative
CC   of MDC, or a nucleic acid encoding MDC or a derivative of MDC. The
CC   products can be used for treating or preventing LV infection or
CC   replication, particularly HIV infection or replication. The products can
CC   also be used for the prognosis for a LV infection, particularly an HIV
CC   infection using the MDC as a prognostic indicator. The methods can also
CC   be used with other LVs, e.g. simian immunodeficiency virus, feline
CC   immunodeficiency virus and bovine immunodeficiency virus.
SQ   Sequence   2923 BP;    605 A;    851 C;    669 G;    788 I;


  Query Match             9.3%;  Score 288.6;  DB 1;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No. 6.4e-55;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy    2573 agaagcactgagctccctgtctctggaagtatttaagaaaaggctgggccaggcacgatg 2632
           | | | |   || | |  |||   | ||   || ||| ||   |||  ||   |||
Db    1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy    2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
           |||||||||||||||| ||| |||||||||  | ||| ||| ||||||| |||||| ||
Db    1805 GCTCACGCCTGTAATCTCAGCACTTTGGGAGGCACAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy    2692 gagtttgagaacagcctggccaacatggtgaaacctcatctctactaaaaatacaaaaat 2751
           |||||  ||| |||||||||||| |||||  | ||||  ||||||||||||||||||||
Db    1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-GTCTCTACTAAAAATACAAAAAT 1687

Qy    2752 tagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgaga 2811
           ||||||| ||| ||||||    |||||||||||||||||||||||||| ||   ||||
Db    1686 TAGCCAGCCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGGGGTGGGAGA 1627

Qy    2812 atcacttaaacctgagaggcagaggttacagtgagccaagatcgtgccactgcattccag 2871
           ||||||| ||||  ||||  ||||||  |||||||| |||||  ||||||||| |||||
Db    1626 AICACTIGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy    2872 cctgggcgacagagcaagactctgtctcaaaaaaaataaaaaat---------------- 2916
           |||||||||||||||  |||| ||| |  | |||||| ||||||
Db    1566 CCTGGGCGACAGAGTGAGACCCTGCCATAAAAAAAAAAAAAAAAGTTGGGGGGGGCAGTGG 1507

Qy    2916 ----------aatcagggcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                     |  | ||||| |||||||| |||||||||||||||||| |||||||||
Db    1506 GGAGAGGGAAGAGGGCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447
```

```
Qy    2965 gaggtgggtggatcacctgaggtcaggagttcgagaccagcctggccaacatggtgaaac 3024
           | |||||||| |||| |||||||  |  |  |  |  |  |  |  |  |  |  |||
Db    1446 GAGGTGGGCGGATTACCTGAGGCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAC 1387

Qy    3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | |||||||| |||| |||||||  |  |  |  |  |  |  |  |  |  |  |||
Db    1386 CTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTAGCAGGAACCTG 1335


RESULT  15
N80616
ID   N80616 standard; DNA; 10475 BP.
AC   N80616;
DT   21-NOV-1990  (first entry)
DE   Genomic clone encoding SAP(Phe).
KW   SAP(Val); SAP(Phe); hyaline membrane disease; ss.
FH   Key             Location/Qualifiers
FT   exon            1054. .1120
FT                   /*tag=  a
FT                   /label=exon I
FT   exon            1431. .1558
FT                   /*tag=  b
FT                   /label=exon II
FT   exon            2060. .2131
FT                   /*tag=  c
FT                   /label=exon III
FT   exon            2495. .2620
FT                   /*tag=  d
FT                   /label=exon IV
FT   exon            3447. .3635
FT                   /*tag=  e
FT                   /label=exon V
FT   exon            3863. .3952
FT                   /*tag=  f
FT                   /label=exon VI
FT   exon            5386. .5568
FT                   /*tag=  g
FT                   /label=exon VII
FT   exon            5742. .5887
FT                   /*tag=  h
FT                   /label=exon VIII
FT   exon            7122. .7202
FT                   /*tag=  i
FT                   /label=exon IX
FT   exon            7697. .7756
FT                   /*tag=  j
FT                   /label=exon X
PN   WO8804324-A.
PD   16-JUN-1988.
PF   03-DEC-1987; U03180.
PR   03-DEC-1987; WO-U03180.
PR   03-DEC-1987; US-939206.
PA   (ABBO) Abbott Laboratories.
PI   Whitsett JA, Fox JL, Pilot-Matias TJ, Meuth JL;
DR   WPI; 88-175472/25.
DR   P-PSDB; P82934.
PI   Pulmonary hydrophobic surfactant-associated protein - useful for
PI   normalising pulmonary surface tension.
PS   Disclosure; pp; English.
CC   This genomic clone encodes the pulmonary hydrophobic surfactant-
CC   associated protein SAP(Phe). Following processing to remove the
CC   introns the prod. is obtd. in a high yield. When SAP(Phe) is
CC   combined with a lipid it can be used to reduce or maintain
CC   normal pulmonary surface tension in the alveoli of animals
CC   (esp. humans) and therefore can be used in the treatment of
CC   hyaline membrane disease in premature infants.
CC   May also be used to deliver substances to respiratory epithelial cells.
CC   See also P80572-96, N80613, N80615, N80617 and N82412-13.
SQ   Sequence   10475 BP;    2416 A;    3063 C;    2440 G;    2556 T;


  Query Match             9.3%;  Score 295.2;  DB 1;  Length 10475;
  Best Local Similarity   80.6%;  Pred. No. 1.1e-54;
```

Matches 379; Conservative 0; Mismatches 58; Indels 23; Gaps 3,

```
Qy    2602 atttaaggaaaaggctgggcccaggcacgatggctcacgcctgtaatcccag-actttggga 2661
           |||| | |||| |  |||||||||| | ||||||||||| |||||||||| |||||||||
Db    8461 ATTTTAAAAAATGTAAGGCCAGGCATGGTGGCTCACGCCGGTAATCCCAGCACTTTGGGA 8520

Qy    2662 ggccgaggcaggcggatcacctgaggtgaggagtttgagaacagcctggccaacatggtg 2721
           |||||||||||||| |||||||||||| ||||||| |||| ||||||||||||||| |
Db    8521 GGCCGAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACTAGCCTGGCCAACATGGCG 8580

Qy    2722 aaacctcatctctactaaaaata-----caaaaattagccaggcgtggtggcaggtgcct 2776
           |||||   ||||||||||||||      ||||||||||   || ||||||||||||||
Db    8581 AAACCCTGTCTCTACTAAAAATATTTTTTAAAAATTAGCTGAGCATGGTGGCAGGTGCCT 8640

Qy    2777 gtaatcccagctacttgggaggctgaggcatgagaatcacttaaacctgagaggcagagg 2836
           |||||||||||||||  ||| ||| ||||| ||||||||||| |||||| |||||  |||
Db    8641 GTAATCCCAGCTACTCAGGACGCTTAGGCAGGAGAATCACTTGAACCTGGGAGGCGAAGG 8700

Qy    2837 ttacagtgagccaagatcgtgccactgcattccagcctgggcgacagagcaagactctgt 2896
           || | ||| ||  |||||||||||||||| || ||||| ||| |||||||| |||||||
Db    8701 TTGCGGTGTGCTGAGATCGTGCCACTGCACTCTAGCCTAGGCAACAGAGCACAACTCTGT 8760

Qy    2897 ctcaaaaaaaataaaaaata-----------------tcagggcacagtggctcatgcc 2939
           ||||  ||||| |||||| ||                 | |||| ||||||||||  ||
Db    8761 CTCAGGAAAAAAAAAAAAAAAAAAAAAAAGGTATTTCTTTGCTGGGCGCAGTGGCTCACACC 8820

Qy    2940 tgtaatcccagcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcaag 2999
           |||||||||||||| |||||| | |||| | |||||||| ||||||||||||||||||
Db    8821 TGTAATCCCAGCACTTTGGGAGACCGAGGCGAGTGGATCACTTGAGGTCAGGAGTTCAAG 8880

Qy    3000 accagcctggtgaacatggcgaaaccccgtctctaataaaaatacaaaaa 3049
           |||||||    ||||||  |||||||||| |||| |||||| | |||||
Db    8881 ACCAGCCTTACCAACATGATGAAACCCCGTATCTACTAAAAAAAAAAAAA 8930
```

Search completed: February 8, 2000, 20:41:59
Job time: 767 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2000, 20.41:59 ; Search time 284.47 Seconds
(without alignments)
978.886 Million cell updates/sec

Title: US-09-174-937-3
Perfect score: 1113
Sequence: 1 atgggaccaccgaagccac... .....atcaggaaacaagcatgtcc 1113

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	231.6	20.8	1386	1	T38228	Apoptosis-regulati
2	231.6	20.8	1386	1	X04120	Human Nip2 encodin
3	82.6	7.4	1227	1	X39873	Gastric cancer ass
4	82.6	7.4	816	1	X39876	Gastric cancer ass
5	43.6	3.9	3044	1	Q83378	Cyclic-GMP stimula
6	43.6	3.9	3044	1	T51113	pHcgs6n cyclic GMP
7	43.6	3.9	3044	1	I67224	Human cGS-PDE cDNA
8	43.6	3.9	3044	1	V36157	Human cyclic-GMP-n
9	43.6	3.9	3044	1	V54762	Human cGS-PDE cDNA
10	43.6	3.9	3044	1	V48205	Cyclic-GMP-stimula
c 11	42.4	3.8	30001	1	T61016	Total DNA sequence
c 12	42.4	3.8	30001	1	X85111	S. aureofaciens DN
c 13	42.2	3.8	15079	1	Q81580	S. clavuligerus cl
14	41.8	3.8	4692	1	V04635	Porcine phosphoino
15	41.8	3.8	4692	1	V74099	Porcine G-protein
16	41.2	3.7	4897	1	Q03259	Pseudorabies virus
17	41.2	3.7	3044	1	Q30184	Human foetal clone
18	41.2	3.7	1495	1	T35134	5126 anther specif
19	41	3.7	1538	1	T63302	A. chrysogenum O-a
20	40.6	3.6	390	1	Q21833	Randomising oligon
21	40.6	3.6	390	1	Q36859	PCR primer for 5'
22	40.2	3.6	1161	1	Q63295	Consensus sequence
c 23	40.2	3.6	137507	1	V19941	KSHV long unique c
c 24	40.2	3.6	32207	1	V73805	KSHV LUR DNA (nucl
25	40	3.6	12588	1	Q63293	Sequence encoding
26	40	3.6	5690	1	Q86902	Rat receptor type-
27	39.8	3.6	1010	1	V99255	DNA encoding an ac
28	39.6	3.6	2795	1	N50036	Partial sequence o
c 29	39.4	3.5	2481	1	I13557	Syngp160mn. Synthe
c 30	39.4	3.5	2481	1	T73952	HIV-1 gp160 synthe
31	39.4	3.5	137507	1	V19941	KSHV long unique c
c 32	39.4	3.5	2481	1	V23291	Synthetic HIV-1 gp
33	39.4	3.5	32207	1	V73805	KSHV LUR DNA (nucl
c 34	39.2	3.5	1953	1	N90631	Sequence including
35	39.2	3.5	4039	1	V05886	Human Tbc-1 gene
36	38.8	3.5	1187	1	N91083	DNA sequence encod
37	38.8	3.5	14926	1	V62156	HSV-2 strain SB5 C
38	38.8	3.5	117213	1	V62179	HSV-2 strain SB5 C
39	38.6	3.5	2277	1	V05372	Human telomerase p
40	38.6	3.5	2277	1	V18886	Homo sapiens mamma
41	38.6	3.5	2634	1	V22394	Microbispora therm
42	38.4	3.5	2271	1	I79277	T. niveum GAPDH. N
43	38.4	3.5	1348	1	I79278	T. niveum GAPDH. N
44	38.4	3.5	2857	1	I79281	T. niveum GAPDH gp
45	38.4	3.5	24379	1	I93195	Streptomyces freno

ALIGNMENTS

RESULT 1
T38228
ID T38228 standard; cDNA; 1386 BP.
AC T38228;
DT 18-DEC-1996 (first entry)
DE Apoptosis-regulating protein Nip2 cDNA.
KW Apoptosis-regulating protein, Nip2; adenovirus E1B 19K protein;
KW Bcl-2; oncogene, cell death; cancer; tumour; immune disorder;
KW diagnosis; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 214. .1158
FT /*tag= a
PN EP-733706-A2.
PD 25-SEP-1996.
PF 21-MAR-1996; 104542.
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 96-427055/43.
DR P-PSDB; W00998.
PT Nucleic acids encoding apoptosis regulating proteins - useful for
PT diagnosing and treating immune disorders, malignancies, etc.
PS Claim 1; Page 39-41; 60pp; English.
CC 6 cDNA clones (T38227-32) respectively, code for novel apoptosis-
CC regulating proteins Nip1, Nip2, Nip3, BipIA, BipF and BipB
CC (W00997-W01002). Nip clones were isolated from human cDNA
CC libraries using a yeast two-hybrid strategy with adenovirus E1B
CC 19K protein (W01010) as bait. Bip clones were isolated using
CC Bcl-2 oncoprotein (W01018) as bait. E1B 19K and Bcl-2 both protect
CC cells against apoptosis. Nip1, Nip2, Nip3 bind to 19K, BipIA and
CC Bip5 bind to 19K and Bcl-2. Nip clones have BH3 domain. Nucleic
CC acids can be used in diagnostic assays, to produce recombinant
CC apoptosis-regulating proteins, or to regulate cell death by
CC increasing or suppressing expression of an apoptosis regulating gene
CC such as bcl-2.
SQ Sequence 1386 BP; 428 A; 249 C; 336 G; 373 T;

Query Match 20.8%; Score 231.6; DB 1; Length 1386;
Best Local Similarity 56.8%; Prod. No. 1.4e-47;
Matches 551; Conservative 0; Mismatches 374; Indels 45; Gaps 5;

Qy 9 cagcgaagccaagctcagcatgaaaacgtgaacgtgaaagaaatgacaaaaacaaaga 68
Db 195 CATTGAATCCAGCCTCACCATCAAACTTTCAAATTAAATAAAAAAAAAAAAAACAACA 254

Qy 69 tcttcccaggccactcccagaagagacgggagtggaactgcttgtcaggcccgatgaaaga 128
Db 255 TTTTCCGATACCTTTACCAGAAGATGATATGTATTGAACCAGATATACTACCTATAAATTGG 314

Qy 129 cagatcctctcctcccaacaccctaaatttcaaccgagaccatcctaaggaaaagaccct 188
Db 315 ACCAGAGGACCAGCCTGGCTCACTACAACTTAATGTAA---ATAAATGAGAAAAAAACT 371

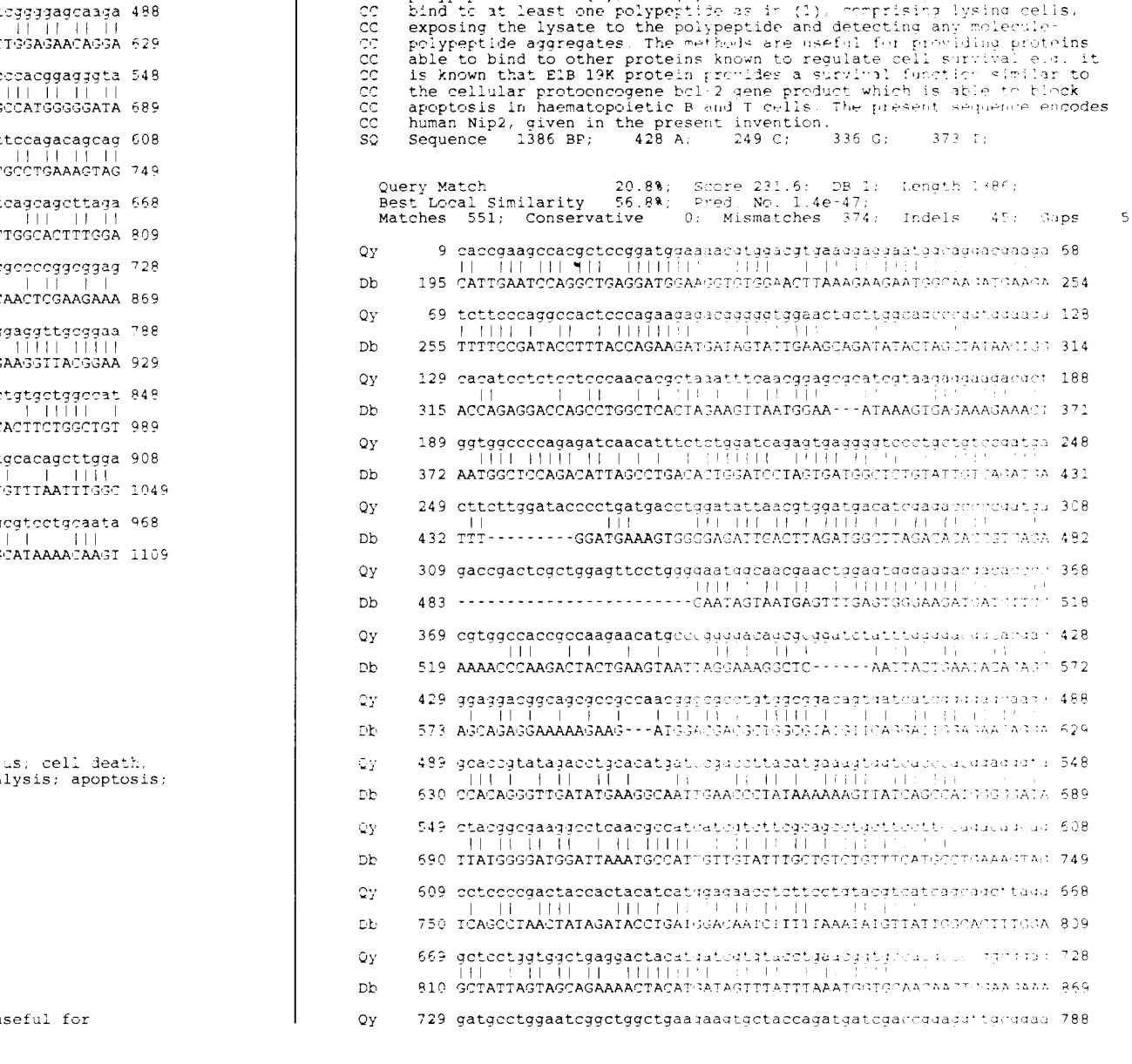
Qy 189 ggtgggcccagagatcaacatttctctggatcagatcaaaggtcctgtgtgtgtcaaga 248
Db 372 AATGGCTCCAGACATTAGCCTCACACTCCATCCTATGTGATGGCTTGTATTGTCAGATGA 431

Qy 249 cttattggatacccctgatgaccttgatattaagtgtgttgacatcgaggaaatcagaaga 308
Db 432 TTT---------GGATGAAAGTGGGAGATTCACTTAGATGGCTAGACACACCATTAGA 482



RESULT 2
X04120
ID X04120 standard; cDNA; 1386 BP.
AC X04120;
DT 13-APR-1999 (first entry)
DE Human Nip2 encoding cDNA.
KW Human; Nip1; Nip2; Nip3; Bip1A, Bip3, Bip13; adenovirus; cell death;
KW viral infection; Bcl-2; protooncogene; mutational analysis; apoptosis;
KW E1B 19K protein; cell survival regulation; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 214..1158
FT /*tag= a
PN US5858678-A.
PD 12-JAN-1999.
PF 21-MAR-1995; 408095.
PR 02-AUG-1994; US-284139.
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 99-152099/13.
DR P-PSDB; W94324.
PT Polypeptides that bind to anti-apoptotic proteins - useful for
PT protecting against cell death induced by viral infection and to
PT modulate response to physical and chemical stimuli
PS Claim 2; Column 51-54; 41pp; English.
CC The present invention describes: (1) a method for regulating cell death,
CC comprising exposing an isolated cell to a polypeptide selected from
CC Nip1, Nip2, Nip3, Bip1A, Bip3 and Bip13, (2) a method for neutralising
CC the activity of the adenovirus E1B 19 kD protein, the Bcl-2 protein or
CC the BHRF-1 protein, comprising exposing an isolated cell to a
CC polypeptide as in (1); and (3) a method for detecting molecules that
CC bind to at least one polypeptide as in (1), comprising lysing cells,
CC exposing the lysate to the polypeptide and detecting any molecule-
CC polypeptide aggregates. The methods are useful for providing proteins
CC able to bind to other proteins known to regulate cell survival e.g. it
CC is known that E1B 19K protein provides a survival function similar to
CC the cellular protooncogene bcl-2 gene product which is able to block
CC apoptosis in haematopoietic B and T cells. The present sequence encodes
CC human Nip2, given in the present invention.
SQ Sequence 1386 BP; 428 A; 249 C; 336 G; 373 T;

Query Match 20.8%; Score 231.6; DB 1; Length 1386;
Best Local Similarity 56.8%; Pred. No. 1.4e-47;
Matches 551; Conservative 0; Mismatches 374; Indels 45; Gaps 5;



```
Db    870 AATGCCCAGTCTGGGATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAA 929

Qy    789 aaacctgaagtccttgatcatcgtccacccctcgtggttcattcggactgtgctggccat 848

Db    930 AAATCTAAAATCCCTAATCATTGTACATCCTTTGGTTTATCAGAACACTTCTGGCTGT 989

Qy    849 ctctcgccctttcatcagcgtcaagtttctctcaacaagatccagtacgtgcacagcttgga 908

Db    990 TACAAGACCATTTATTAGCTCGAAATTCAGCTAAAAAATTAGATACGTGTTTAATTTGGC 1049

Qy    909 agacctgaagcaactcatcccatatggaacagtccagatcccagactgcgtcctgcaata 968

Db   1050 AGAACTAGCAGAACTTGTCCCCATGGAATACTTTGGCATACTAGAATGCATAAAACAAGT 1109

Qy    969 cgaagaggaa 979

Db   1110 TGATCAAGAA 1119


RESULT   3
X39873
ID   X39873 standard; DNA; 1227 BP.
AC   X39873;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; 98IB14679.
PR   22-JUN-1998; US-1002322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132449/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 579-580; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   1227 BP;    398 A;    270 C;    304 G;    254 T;


  Query Match             7.4%;  Score 82.6;  DB 1;  Length 1227;
  Best Local Similarity   62.0%;  Pred. No. 2.6e-11;
  Matches  147;  Conservative    0;  Mismatches   89;  Indels    1;  Gaps    1;

Qy     97 qqqgtggaactgcttggcagccccggatggaagacacatcctctcctcccaacacgctaaat 156

Db    921 GAGGGCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGAT 980

Qy    157 ttcaacggaaggcatcgtaaga                                        216

Db    981 CTTAATGACACTCATCCTCGGA                                        1040

Qy    217 gatcagagtgaaggggtcccctgctctgatgatgttctttttt                   275

Db   1041 GACCAAAGTGAAGGATCTATTC                                        1100

Qy    276 tattaacatggatatacatggaaa                                     312

Db   1101 CATCAATGTGGATGAACTTGATA                                      1137


RESULT   4
X39876
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; 98IB14679.
PR   22-JUN-1998; US-1002322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132449/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    219 G;    175 T;


  Query Match             7.4%;  Score 82.6;  DB 1;  Length 816;
  Best Local Similarity   62.0%;  Pred. No. 2.3e-11;
  Matches  147;  Conservative    0;  Mismatches   89;  Indels    1;  Gaps    1;

Qy     97 qqqgtggaactgcttggcagccccggatggaagacacatcctctcctcccaacacgctaaat 156

Db    510 GAGGGCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGAT 569

Qy    157 ttcaacggaaggcatcgtaaga                                        216

Db    570 CTTAATGACACTCATCCTCGGA                                        629

Qy    217 gatcagagtgaaggggtcccctgctctgatgatgttctttttt                   275

Db    630 GACCAAAGTGAAGGATCTATTC                                       689
```

```
Qy    276 tattaacatggatgacatcaagaccccggatagaccgactcgctggaagttcctggg 332
          ||  ||  |||||||  | ||  |||||||||    | || ||  | ||||  ||  ||
Db    690 CAICAAIGIGGAIGAACIIGAIACCCCCGAIGAAGCAGAIICITTTGAGIACCCIGG 746


RESULT   5
Q83978
ID   Q83978 standard; cDNA; 3044 BP.
AC   Q83978;
DT   11-OCT-1995  (first entry)
DE   Cyclic-GMP stimulated nucleotide PDE clone HFB9.2.
KW   Cyclic-GMP stimulated nucleotide phospho-diesterase; antibodies;
KW   human; hormones; neurotransmitters; transmission regulation;
KW   enzyme purification; clone HFB9.2; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             12. .2837
FT                   /*tag= a
PN   US5389527-A.
PD   14-FEB-1995.
PF   19-APR-1991; 688356.
PR   19-APR-1991; US-688356.
PR   20-APR-1992; US-872644.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA,  Charbonneau H,  Sonnenburg WK;
DR   WPI; 95-090205/12.
DR   P-PSDB; R69729.
PT   New nucleic acid encoding cyclic-GMP stimulated nucleotide
PT   phospho-di.esterase - and related vectors and transformed cells,
PT   useful for screening cpds. for phospho-di.esterase modulating
PT   activity
PS   Claim 1; Columns 89-96; 69pp; English.
CC   Qes478 encodes Ke272s the human cyclic-GMP stimulated nucleotide
CC   phospho-diesterase (CaM PDE) clone HFB9.2  Eukaryotic cells
CC   that express CaM PDE can be used to screen cpds. for the ability to
CC   modulate CaM PDE activity  CaM PDEs are involved in regulating
CC   the transmission of information from hormones, neurotransmitters
CC   or other systems that use cyclic nucleotides as messengers.
CC   Antibodies raised against CaM PDE can be used for enzyme purifcn.,
CC   or determination.
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;

  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtgagggtccctgctatccgatgagttct 253
          ||| || | |   |||  | || |  | |  |||                || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccccctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccg 313
          |||    |  ||   | |||  |  ||| |   |    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actcgctggaagttcctggggaatggcaacgaactggagtgggaagacgacaccccgtgg 373
          |  | ||||    |||| |||  |   | |   || ||  |   |    ||     |
Db    559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCCTGCAGCAGCGCGGGCCCAGGAGGCTC 618

Qy    374 ccaccgccaagaacatgcccgggacagcgcggaatctatttggggacggcacgacggagg 433
          ||   |||    | |  |||  |||  |  ||      |  | |||  | |||  | |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGCAGCGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcgggggagcaagagcacc 493
          |  |  | ||  |  | |  |||    |  |  |  ||  || |  ||   |
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |  |  | || ||  | |  |  ||  |  |   ||  ||||  | |    || |
Db    739 CTCCCIGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT   6
T51113
ID   T51113 standard; cDNA; 3044 BP.
AC   T51113;
DT   17-MAR-1997  (first entry)
DE   pHcgs6n cyclic GMP stimulated phosphodiesterase cDNA.
KW   Cyclic GMP stimulated phosphodiesterase; cGS-PDE; bovine; brain; cAMP;
KW   adrenal gland, cGMP, transmembrane signal; extracellular hormone;
KW   neurotransmitter; antibody; human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             12. .2837
FT                   /*tag= a
FT                   /product= cyclic GMP stimulated phosphodiesterase
PN   US5580771-A.
PD   03-DEC-1996.
PF   19-APR-1991; 688356
PR   19-APR-1991; US-688356.
PR   20-APR-1992; US-872644.
PP   29-AUG-1994; US-297494.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA,  Charbonneau H,  Sonnenburg WK;
DR   WPI; 97-033573/03.
DR   P-PSDB; W11253.
PT   DNA encoding cGMP-stimulated phospho-di esterase - for prodn. of
PT   recombinant enzyme
PS   Example 6; Column 89-96; 68pp; English.
CC   This sequence represents the coding sequence for the cyclic GMP
CC   stimulated phosphodiesterase (cGS-PDE, clone pHcgs6n isolated from human
CC   foetal brain.  The cyclic nucleotide phosphodiesterases (PDEs) catalyse
CC   the hydrolysis of 3', 5' cyclic nucleotides such as cAMP and cGMP, to
CC   their corresponding 5'-nucleotide monophosphates.  The PDEs are therefore
CC   important in the control of the cellular concentration of cyclic
CC   nucleotides.  The PDEs are, in turn, regulated by transmembrane signals
CC   or second messenger ligands such as calcium ion or cGMP.  The PDEs
CC   therefore have a central role in regulating the flow of information from
CC   extracellular hormones, neurotransmitters, or other signals that use the
CC   cyclic nucleotides as messengers.  PDEs are present in most of the cells
CC   and tissues of eukaryotic organisms, but only in trace amounts.  cGS-PDEs
CC   are stimulated by cGMP in effecting cAMP hydrolysis.  This is thought to
CC   occur through a noncatalytic cGMP-specific site present on the cGS-PDE.
CC   This sequence can be used for the production of recombinant cGS-PDE,
CC   which may have therapeutic and diagnostic uses.  The encoded protein may
CC   also be useful for diagnostic antibody production.
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;

  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtgagggtccctgctatccgatga[illegible] 253
          ||| || | |   |||  | || |  | |  |||                || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccccctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccg 313
          |||    |  ||   | |||  |  ||| |   |    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actcgctggaagttcctggggaatggcaacgaactggagtgggaagacgacaccccgtgg 373
          |  | ||||    |||| |||  |   | |   || ||  |   |    ||     |
Db    559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCCTGCAGCAGCGCGGGCCCAGGAGGCTC 618

Qy    374 ccaccgccaagaacatgcccgggacagcgcggaatctatttggggacggcacgacggagg 433
          ||   |||    | |  |||  |||  |  ||      |  | |||  | |||  | |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGCAGCGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcgggggagcaagagcacc 493
          ||  |  | ||  |  | |  |||    |  |  |  ||  || |  ||   |
```

```
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |       | || ||   | | |  ||| | |  |   ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT   7
I67224
ID   I67224 standard; cDNA; 3044 BP.
AC   I67224;
DT   05-AUG-1997  (first entry)
DE   Human cGS-PDE cDNA sequence from pHcgs6n.
KW   Cyclic-GMP-stimulated phosphodiesterase; antibody; human;
KW   polymerase chain reaction; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             12. .2834
FT                   /*tag= a
FT                   /product= cGS-PDE
PN   US5602019-A.
PD   11-FEB-1997.
PF   19-APR-1991; 688356.
PR   19-APR-1991; US-688356.
PR   20-APR-1992; US-871044.
PR   23-AUG-1994; US-297510.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR   WPI; 97-131799/12.
DR   P-PSDB; W18050.
PT   DNA encoding bovine and human phosphodiesterase enzymes - stimulated
PT   by calcium/calmodulin, useful for recombinant prodn. of the enzymes
PS   Example 6; Column 91-98, 69pp, English.
CC   The present sequence represents the human cyclic-GMP-stimulated
CC   phosphodiesterase cGS-PDE from pHcgs6n cDNA sequence. The DNA
CC   sequences of the cGS-PDE's are used for the production of the
CC   recombinant enzymes, which in turn may be used for antibody
CC   production and to screen for compounds that modulate
CC   phosphodiesterase activity.
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 cccagagatcaacatttctctggatcagagtgagggtccctgctgtccgatgacttct 253
          ||| || | |   ||| | ||  |  | |   |||          || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccctgatgaccttggatattaacgtggatgacatcgagaccccccgatgagaccg 313
          |||   |  ||  | |||  |  |||   ||    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actgctggagttcctgggaaatggcaaccaacctggagtggtaagacgacaccccctgtgg 373
          | | ||||    |||| |||  |  | |  ||| ||  |  |     ||        |
Db    559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCTTGCAGCAGCGCGGGCCCAGGGAGCTC 618

Qy    374 ccaccgccaagaacatgccccgggacagccgcggatctatttggggacggcacgacggagg 433
          ||   |||     | |  |||  |||  |  ||         |  | |||  | |||  |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACCGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcgggagcaagagcacc 493
          ||  |  |  || |  |  |||     ||  |     ||   || | |  ||    |
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |  |  || ||  |  | |  |||  ||  |     ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT   8
V36157
ID   V36157 standard; cDNA; 3044 BP.
AC   V36157;
DT   07-SEP-1998  (first entry)
DE   Human cyclic-GMP-nucleotide phosphodiesterase cDNA.
KW   Bovine; decrease; intracellular concentration; cAMP; cGMP;
KW   calcium ion/calmodulin stimulated cyclic nucleotide phosphodiesterase;
KW   CaM-PDE; therapy; prognosis; diagnosis; specific modulator;
KW   CaM-PDE activity; identification; signalling; extracellular hormone;
KW   neurotransmitter; selective inhibitor; cardiotonic; antidepressant;
KW   antihypertensive; antithrombotic agent;
KW   cyclic-GMP-nucleotide phosphodiesterase; cGS-PDE; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             12. .2837
FT                   /*tag= a
FT                   /product= cGS-PDE
PN   US5776752-A.
PD   07-JUL-1998.
PF   07-JUN-1995; 473532.
PR   20-APR-1992; US-872044.
PR   19-APR-1991; US-688356.
PR   23-AUG-1994; US-297494.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR   WPI; 98-398038/34.
DR   P-PSDB; W60754.
PT   New calmodulin and calcium stimulated cyclic nucleotide
PT   phosphodiesterase - used e.g. to identify selective modulators,
PT   potentially useful as e.g. cardiotonic, antidepressant,
PT   antihypertensive agents
PS   Example 6; Columns 89-96; 69pp; English.
CC   The present sequence encodes human cyclic-GMP-nucleotide
CC   phosphodiesterase (cGS-PDE). It was isolated using bovine adrenal cortex
CC   cGS-PDEs. The specification describes methods for identifying stimulated
CC   cyclic nucleotide phosphodiesterases (CaM-PDEs). CaM-PDEs are
CC   characterised by their responsiveness to intracellular calcium, which
CC   leads to a decreased intracellular concentration of cAMP and/or cGMP.
CC   CaM-PDEs, and their fragments or synthetic peptides, are used in therapy,
CC   prognosis and diagnosis. They are also used to generate specific
CC   antibodies useful as immunoassay reagents for detecting the protein or
CC   for affinity purification and for identifying specific modulators of
CC   CaM-PDE activity. CaM-PDEs are implicated in signalling involving
CC   extracellular hormones and neurotransmitters, so selective inhibitors are
CC   potentially useful as cardiotonic, antidepressant, antihypertensive and
CC   antithrombotic agents.
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 cccagagatcaacatttctctggatcagagtgagggtccctgctgtccgatgacttct 253
          ||| || | |   ||| | ||  |  | |   |||          || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccctgatgaccttggatattaacgtggatgacatcgagaccccccgatgagaccg 313
          |||   |  ||  | |||  |  |||   ||    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actgctggagttcctgggaaatggcaaccaacctggagtggtaagacgacaccccctgtgg 373
          | | ||||    |||| |||  |  | |  ||| ||  |  |     ||        |
Db    559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCTTGCAGCAGCGCGGGCCCAGGGAGCTC 618

Qy    374 ccaccgccaagaacatgccccgggacagccgcggatctatttggggacggcacgacggagg 433
          ||   |||     | |  |||  |||  |  ||         |  | |||  | |||  |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACCGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcgggagcaagagcacc 493
          ||  |  |  || |  |  |||     ||  |     ||   || | |  ||    |
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |  |  || ||  |  | |  |||  ||  |     ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800
```

```
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACGGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaagagcacc 493
          || |  |  || |  |  |  |||    | |   | |    ||  || |  ||   |
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |      | || ||   | |  |  ||| || |   ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT   9
V54762
ID   V54762 standard; cDNA; 3044 BP.
AC   V54762;
DT   29-OCT-1998  (first entry)
DE   Human cGS-PDE cDNA DNA sequence.
KW   Ca2+/calmodulin sensitive cyclic nucleotide phosphodiesterase; CaM-PDE,
KW   assay; identification; modification; enzymatic activity; modulator;
KW   cyclic guanosine monophosphate-stimulated phosphodiesterase; cGS-PDE;
KW   human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             12. .2837
FT                   /*tag=  a
FT                   /product= cGS-PDE
PN   US5800987-A.
PD   01-SEP-1998.
PF   31-MAY-1995; 455525.
PR   20-APR-1992; US-872644.
PR   19-APR-1991; US-688356.
PR   29-AUG-1994; US-297494
PR   31-MAY-1995; US-455525.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR   WPI; 98-494762/42.
DR   P-PSDB; W71226.
PT   Identification of modulators of Calcium/calmodulin sensitive cyclic
PT   nucleotide phosphodiesterase - uses recombinant cells expressing the
PT   enzyme and monitoring protein expression in the presence of
PT   potential modulators
PS   Example 6; Columns 91-98; 69pp; English.
CC   The present sequence encodes human cyclic guanosine
CC   monophosphate-stimulated phosphodiesterase (cGS-PDE) cDNA. The
CC   specification also describes Ca2+/calmodulin sensitive cyclic
CC   nucleotide phosphodiesterase (Cam-PDE) enzymes. CaM-PDE enzymes
CC   catalyse the hydrolysis of cyclic nucleotides and as such control their
CC   intracellular level. They are controlled by second messengers e.g. Ca2+
CC   and calmodulin and transmembrane signals and through pathways involving
CC   these, regulate flow of information from extracellular hormones,
CC   neurotransmitters and other signals using cyclic nucleotides as
CC   messengers. The specification describes an assay for identifying a
CC   chemical agent which modifies the enzymatic activity of a mammalian
CC   CaM-PDE. The assays are useful for identifying modulators of
CC   CaM-PDEs. The assay is carried out by incubating cells expressing
CC   CaM-PDE with the suspected modulator and measuring its effect
CC   e.g. monitoring the hydrolysis of cAMP and/or cGMP
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtgaggggtccctgctgtccgatgacttct 253
          ||| || | |   ||| | || |   | |  ||||         || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccctgatgacctggatattaacgtggatgacatcgaggaccccgatgaagaccg 313
          |||    |  ||   | |||  |  ||| | |||    | |  |    | |  |||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAAAGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actcgctggagttcctggggaatggcaacgaactggagtggaagacgacaccccctgg 373
          |  | ||||    |||| |||  |   | |  ||| ||  |   |    ||     |
Db    559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCCTGCAGCAGCGCGGGCCCAGGAGGCTC 618

Qy    374 ccaccgccaagaacatgcccgggacacgcgcggatctatttggaggacaggcacgacgaagg 433
          ||  |||   | |  |||  |||  |  |||    |     | | |||  | |||  |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACGGCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy    434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaagagcacc 493
          ||  |  |  || |  |  |  |||    | |   | |    ||  || |  ||   |
Db    679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy    494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
          |      | || ||   | |  |  ||| || |   ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT  10
V48205
ID   V48205 standard; cDNA; 3044 BP.
AC   V48205;
DT   14-DEC-1998  (first entry)
DE   Cyclic-GMP-stimulated phosphodiesterase clone pHcgs6n.
KW   ss; cow; Ca2+ stimulated phosphodiesterase;
KW   Calmodulin stimulated phosphodiesterase; therapeutic; antibody.
OS   Bos sp.
FH   Key             Location/Qualifiers
FT   CDS             12. .2837
FT                   /*tag=  a
FT                   /product= "Cyclic-GMP-stimulated phosphodiesterase"
PN   US5789553-A.
PD   04-AUG-1998.
PF   31-MAY-1995; 455526.
PR   20-APR-1992; US-872644.
PR   19-APR-1991; US-688356.
PR   29-AUG-1994; US-297494.
PR   31-MAY-1995; US-455526.
PA   (UNIW ) UNIV WASHINGTON.
PI   Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR   WPI; 98-446185/38.
DR   P-PSDB; W77042.
PT   Antibody to phosphodiesterase polypeptide - useful for purification
PT   or detection of polypeptide
PS   Example 6; Column 89/96; 68pp; English.
CC   The cyclic-GMP-stimulated phosphodiesterase clone pHcgs6n encodes a
CC   phosphodiesterase from the bovine brain. The phosphodiesterases isolated
CC   in this invention were used to transform host cells, preferably mammalian
CC   or yeast cells. The recombinant proteins and fragments produced can be
CC   used for therapeutic, diagnostic, and prognostic purposes and will
CC   provide the basis for preparation of monoclonal and polyclonal
CC   antibodies. An antibody specific for the Ca2+, calmodulin-stimulated
CC   cyclic nucleotide phosphodiesterase (PDE) polypeptide is useful for
CC   affinity purification or detection of the polypeptide and is selective
CC   only for the Ca2+/calmodulin stimulated cyclic nucleotide PDE
CC   polypeptide.
SQ   Sequence   3044 BP;    701 A;    901 C;    846 G;    596 T;


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.11;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtgaggggtccctgctgtccgatgacttct 253
          ||| || | |   ||| | || |   | |  ||||         || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498
```

```
Qy     204 tggatacccctgatgaccctggatattaaagtggatgacatcgagacccccgatgagaccg 313
           |||     |  ||   | |||  |  ||| | |||     | |  |     | |  ||||
Db    4449 TGGTGCACTCTGGCCAGCTGAGTGATAAAGAGAAATGGAGCCTGCAGGCGGTGGAGAAGC 4508

Qy     314 actgctggagttcctgggaatggcaacgacctggagtgggaagacgacacccccgtgg 373
           |  | ||||    |||| |||  |   | |   || ||  |   |    ||     |
Db     4509 ATACCCTGCTCGCCCTGCGGACGGTGCAGCTCCTGCAGCAGTGCGGGCCCAGGGAGGCTC 4568

Qy     374 ccaccgcccaagaacatgcccggggacagccgccggatctatttggggacggcacgacggagg 433
           ||   |||     | |  |||  |||  |  | ||     |     | | |||  | ||| |
Db     4569 CCCGACCCGTCCAGAACCCCCCGGAGGGACTCGCGGAAGACCAGAAGGGCGGGGCGGCGT 4628

Qy     434 acggcagccgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaagagcacc 493
           ||  |  |  ||  |  |  |||     | |    ||   ||  |  ||  |   |
Db     4629 ACACCGACCCCGACCGCAAGATCCTCCAACTCTGCGGGGAACTCTACGACCTGGATGCCT 4688

Qy     494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
           |    | ||  ||   | |  |  ||| ||  |   ||  ||||  | |     || |
Db     4689 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 4748

Qy     554 gc 555
           ||
Db     4749 GC 4800


RESULT  11
T61016/c
ID   T61016 standard; DNA; 30001 BP.
AC   T61016;
DT   21-APR-1997  (first entry)
DE   Total DNA sequence from cosmid clones LP(2)127 and LP(2)128.
KW   Cosmid clone, LP(2)127, LP(2)128, chlortetracycline,
KW   biosynthetic pathway, recombinant, production, antibiotic,
KW   heterologous host, Streptomyces lividans, ss
OS   Streptomyces aureofaciens.
PN   US5593885-A.
PD   31-DEC-1996.
PF   26-JUL-1990; 558039.
PR   26-JUL-1990; US-558039.
PR   26-JUL-1990; US-558040.
PR   15-JAN-1992; US-821419.
PR   15-JAN-1992; US-821109.
PR   22-SEP-1993; US-125468.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Fantini SE, Lotvin JA, Ryan MJ, Strathy N;
DR   WPI; 97-078653/07.
PT   DNA encoding tetracycline biosynthetic pathway proteins -
PT   specifically from Streptomyces aureofaciens for expression in
PT   heterologous hosts, specifically S. lividans
PS   Claim 3; Columns 17-44, 39pp, English.
CC   The present sequence is the total DNA sequence from cosmid clones
CC   LP(2)127 and LP(2)128, which contains a Streptomyces aureofaciens
CC   (AICC 13899) DNA that encodes the proteins of the entire
CC   chlortetracycline biosynthetic pathway. The biosynthetic gene,
CC   which can be expressed in heterologous hosts, especially
CC   S. lividans, may be useful in the production of antibiotics.
SQ   Sequence   30001 BP;    4055 A;    10939 C;    11019 G;    3988 T;


  Query Match             3.8%;  Score 42.4;  DB 1;  Length 30001;
  Best Local Similarity   44.6%;  Pred. No. 0.42;
  Matches  166;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy     360 cgacacccccgtggccaccgccaagaacatgcccggggacagcgcggatctatttggggga 419
           ||  | |    |||||| |||| ||      | || ||  | |    |  ||
Db    8338 CGGCATCGGACTGGCCAGCGCCGAGCGGCTGGCCCGGCGCGGGCTGAGCGTCTTCATCTG 8279

Qy     420 cggcacgaccggaggacggcagcgccgccaccgtgcgcgctgtggcggacagtgatcatcggg 479
           ||  |   || | | | |||  ||||| |  |    ||    |||    | |  | | |
Db    8278 CGCCCGCACCGCCGACGGCGTCGCCGAGACGTTCGCCCGCCTGCTGCGCGGCGAGGGCCTGGA 8219

Qy     480 ggagcaagagcaccgtatagacctgcacatgatccggccttacatgaaagtggtcaccca 539
           |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    8218 GGTGCACGGCCTCGGCGCCCACGTCGCCTTCGCGCTCGACGTCAACGTCGTCGTCGCG 8159

Qy     540 cggaggtactacggcgaaggc 561
           |  |  | |  | |  |  |
Db    8158 GCCGCTGCACCGTTCGGCCG 8137

Qy     600 agacagcaagcctcccgactacgcatcatgtatggcgaaacctctgtgtacgtcatcac 659
           |  | || |  | ||  | |  |  |  |  |  |  |  |  |  |  |  |  |
Db    8099 TGGCCTCACCTCCACCCTCACCTACGGCTGCGGCAGGACGTGACTGACACTAACCTCAC 8039

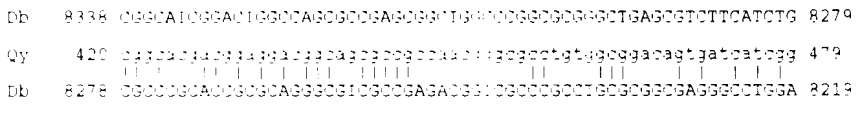
Qy     660 cagcttagagctcctggtggctgaggagctacatgatcgtgtacctaaaaatctccacgtc 719
           ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    8038 CAGCGTCTTCCGCGTCACCCGCCAGCTCCTCACCACCCCCGCCATCGACCCCGCCGATA 7979

Qy     720 ccgcggaggat 731
           | ||||||  ||
Db    7978 CGGGCGGATCAT 7967

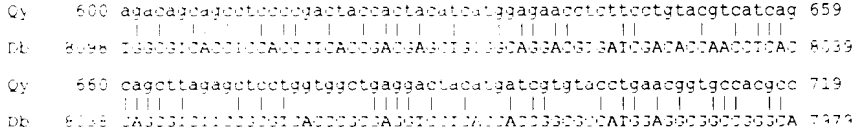

RESULT  12
X05110/c
ID   X05110 standard; DNA; 30001 BP.
AC   X05110;
DT   07-APR-1999  (first entry)
DE   S. aureofaciens DNA in cosmid clones LP2-127 and LP2-128.
KW   Chlortetracycline; tetracycline; biosynthetic; actinomycete;
KW   S. ariseofuscus, S. ambofaciens; antibiotic resistance gene;
KW   S lividans; S aureofaciens, cosmid clone, antibiotic, ss.
OS   Streptomyces aureofaciens.
PN   US5866410-A.
PD   02-FEB-1999.
PF   07-JUN-1995; 474933.
PF   22-SEP-1993; US-125468.
PF   26-JUL-1990; US-558039.
PF   26-JUL-1990; US-558040.
PR   15-JAN-1992; US-821109.
PR   15-JAN-1992; US-821419.
PR   07-JUN-1995; US-474933.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Fantini SE, Lotvin JA, Ryan MJ, Strathy N;
DR   WPI; 99-141936/12.
PT   Two plasmids for cloning the biosynthetic pathways of tetracycline,
PT   chlortetracycline, and their derivatives - comprise an origin of
PT   replication, cohesive end sites and optionally, an antibiotic
PT   resistance gene
PS   Disclosure; Fig 4A-L; 39pp; English.
CC   The invention relates to combination of two plasmids for cloning the
CC   genes encoding the biosynthetic production pathways of chlortetracycline,
CC   tetracycline, or their derivatives. One plasmid contains an origin of
CC   replication, an actinomycete (such as Streptomyces lividans,
CC   S. ariseofuscus, or S. ambofaciens) active antibiotic resistance gene and
CC   three or more tandem cohesive end sites. The second plasmid contains an
CC   actinomycete active origin of replication and three or more tandem
CC   cohesive end sites. The invention can be used to clone a large amount of
CC   genetic material for the heterologous production of the antibiotics,
CC   chlortetracycline, tetracycline and their analogues. A cluster of genes
CC   can be inserted into the plasmid allowing a biosynthetic pathway to be
CC   transferred in its entirety to a heterologous host. The invention
CC   specifically relates to cloning of the entire tetracycline and
CC   chlortetracycline from Streptomyces aureofaciens and its expression in a
CC   heterologous host such as S. lividans. The present sequence represents a
CC   S. aureofaciens DNA sequence from the cosmid clones designated LP2-127
CC   and LP2-128.
SQ   Sequence   30001 BP;    4055 A;    10939 C;    11019 G;    3988 T;


  Query Match             3.8%;  Score 42.4;  DB 1;  Length 30001;
  Best Local Similarity   44.6%;  Pred. No. 0.42;
  Matches  166;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy     360 cgacacccccgtggccaccgccaagaacatgcccggggacagcgcggatctatttggggga 419
           ||  | |    |||||| |||| ||      | || ||  | |    |  ||
```



RESULT 13
Q91580/c
ID Q91580 standard; DNA; 15079 BP.
AC Q91580;
DT 13-NOV-1995 (first entry)
DE S. clavuligerus cla gene region.
KW Clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;
KW cla gene; ds.
OS Streptomyces clavuligerus.
FH Key Location/Qualifiers
FT cds complement (49. .1745)
FT /*tag= a
FT /label= ORF1
FT cds 2216. .3937
FT /*tag= b
FT /label= ORF2
FT cds 3940. .5481
FT /*tag= c
FT /label= ORF3
FT cds 5654. .6595
FT /*tag= d
FT /label= ORF4
FT /note= "cla gene"
FT cds 6611. .7588
FT /*tag= e
FT /label= ORF5
FT cds 7895. .9076
FT /*tag= f
FT /label= ORF6
FT cds 9241. .10908
FT /*tag= g
FT /label= ORF7
FT cds complement (10998. ..12[illegible]
FT /*tag= h
FT /label= ORF8
FT cds complement (12662. .13865)
FT /*tag= i
FT /label= ORF9
FT cds 13769. .14995
FT /*tag= j
FT /label= ORF10
PN CA2108113-A.
PD 09-APR-1995.
PF 08-OCT-1993; 108113.
PR 08-OCT-1993; CA-108113.
PA (UYAL-) UNIV ALBERTA.
PI Aidoo KA, Jensen SE, Paradkar AS;
DR WPI; 95-207301/28.
DR P-PSDB; P77858; P77859; P77860; P77861; P77862; P77863; P77864;
DR P77865; P77866; P77867
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
PT biosynthesis of the antibiotic in Streptomyces hosts which do not
PT naturally produce clavulanate
PS Disclosure; Fig.2; 41pp; English.
CC A library of S. clavuligerus NRRL 3741 genomic DNA fragments (15-22
CC kb) constructed in cosmid pLAFR3 was screened for the cla gene using
CC a probe based on a partial N-terminal sequence from the CLA enzyme.
CC Isolated clone K6L1 included a 15kb fragment having the sequence
CC given in Q91580 that included the cla gene (ORF4).
SQ Sequence 15079 BP; 2097 A; 5488 C; 5324 G; 2170 T;



RESULT 14
V04633
ID V04633 standard; cDNA; 4692 BP.
AC V04633;
DT 17-AUG-1998 (first entry)
DE Porcine phosphoinositide 3OH-kinase p101 subunit cDNA.
KW G-beta-gamma regulated phosphatidylinositol-3' kinase; pik;
KW phosphoinositide 3OH-kinase; PI3K; signal transduction;
KW phosphatidylinositol (3,4,5)-tr phosphate; G-protein; receptor;
KW transgenic animal; knockout animal; inflammation; arthritis;
KW septic shock; adult respiratory distress syndrome; pneumonia;
KW asthma; allergy; reperfusion injury; atherosclerosis; cancer;
KW Alzheimer's disease; cancer; antisense; ribozyme; diagnosis;
KW therapy; drug screening; ss.
OS Sus scrofa.
FH Key Location/Qualifiers
FT CDS 390. .3023
FT /*tag= a
PN WO9749818-A2.
PD 31-DEC-1997.
PF 26-JUN-1997; U11219.
PR 27-JUN-1996; US-672211.
PA (ONYX-) ONYX PHARM.
PI Braselmann S, Hawkins PT, Stephens L,

```
DR   WPI; 98-077181/07.
DR   P-PSDB; W23946.
PT   DNA encoding G-beta-gamma regulated phosphatidyl-inositol-3' kinase,
PT   p101 and p120 subunits - useful for diagnosis, drug screening,
PT   clinical trial monitoring and treatment of inflammatory disorders
PS   Claim 2; Fig 1A-D; 151pp; English.
CC   This cDNA sequence codes for the p101 regulatory subunit (see
CC   W23946) of pig G-protein regulated phosphatidylinositol-3' kinase,
CC   a heterodimeric enzyme which produces the intracellular messenger
CC   phosphatidylinositol (3,4,5)-triphosphate in response to activation
CC   of trimeric G protein-linked receptors. This novel protein, which
CC   also contains a catalytic subunit, p120 (see V04634), is found in
CC   cells of haematopoietic origin and is involved in immune system
CC   responses which cause inflammation. The p101 subunit is largely
CC   responsible for the dramatic stimulation of kinase activity in the
CC   presence of activated trimeric G proteins. p101 cDNA was obtained
CC   from a pig neutrophil cDNA library using a degenerate oligonucleotide
CC   probe (see T99714) based on an isolated p101 tryptic peptide. The
CC   p101 cDNA clone in pCMV3mycp101 is deposited as ATCC 97636. The
CC   invention encompasses pig and human p101 and p120 nucleotides,
CC   host cell expression systems, p101 and p120 proteins (see
CC   W23946-49), fusion proteins, polypeptides and peptides, antibodies
CC   to these proteins, and transgenic animals and knockout animals.
CC   Compounds which are useful for treating inflammatory response
CC   disorders can be identified by screening assays using a G protein
CC   activated PI3K, or a cultured host cell that expresses the p101
CC   gene. Antagonists of G protein stimulated PI3K (acting through the
CC   p101 subunit, especially by disrupting the interaction between the
CC   p101 and p120 subunits) can be used to treat arthritis, septic
CC   shock, adult respiratory distress syndrome (ARDS), pneumonia,
CC   asthma, allergies, reperfusion injury, atherosclerosis, cancer and
CC   Alzheimer's disease. Signalling-incompetent p101 (preferably a
CC   deletion mutant) oligonucleotide constructs, antisense
CC   oligonucleotides, triple helix forming oligonucleotides, ribozymes
CC   or recombinant DNA constructs that act through targeted homologous
CC   recombination are all useful for inhibiting expression of the p101
CC   subunit, especially by delivery to cells of haematopoietic origin.
CC   (All claimed). The nucleic acids and their products can also be
CC   used for diagnosis, drug screening and clinical trial monitoring of
CC   inflammatory diseases.
SQ   Sequence   4692 BP;    916 A;    1559 C;    1394 G;    823 T;

  Query Match              3.8%;  Score 41.8;  DB 1;  Length 4692;
  Best Local Similarity   55.1%;  Pred. No. 0.33;
  Matches  102;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy     855 ccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttggaagacct 914
           ||| ||     || || || |  || | | | |  | ||  | |||||||| ||   |||
Db    1232 CCCCATCCCCGTCGCCAGGTGCTACACCTCACCTCGTGAACCAGGACAGCTTCGACATCCT 1291

Qy     915 ggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgcaatacgaaga 974
           | || || || | |  | ||| || |  | | ||          |||| |  |||| ||
Db    1292 GTAGGAAATCCTGCTCAAGGAGTAGGAGCTCCTCCAGCCAGAGATCCTGGACGACGAGGA 1351

Qy     975 ggaa-agactgaaggccaggagggagagcgcgaggccccagccggagtttgtgctgccca 1033
           |||  ||   || |   | |||| ||||  |  || | | | |||    || || |
Db    1352 GGACGAGGACGAGGAGGACGAGGAAGAGGACTTGGACGCCGACGGCCACTGCGCGGAGAG 1411

Qy    1034 ggtct 1038
           || ||
Db    1412 GGACT 1416


RESULT  15
V74099
ID   V74099 standard; DNA; 4692 BP.
AC   V74099;
DT   09-MAR-1999  (first entry)
DE   Porcine G-protein regulated PI3K p101 adapter subunit DNA.
KW   Phosphoinositide 3-hydroxykinase; PI3K; trimeric G protein; porcine;
KW   adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
KW   detection, diagnosis, activation disorder; haematopoietic system;
KW   treatment; immune disorder; inflammation; arthritis; septic shock;
KW   adult respiratory distress syndrome; pneumonia; asthma; allergy;
KW   reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.
OS   Sus scrofa.
PN   US5856132-A.
PD   05-JAN-1999.
PF   15-AUG-1997; 916917.
PR   15-AUG-1997; US-916917.
PR   27-JUN-1996; US-672211.
PA   (ONYX-) ONYX PHARM.
PI   Braselmann S, Hawkins PT, Stephens L;
DR   WPI; 99-105107/09.
DR   P-PSDB; W90082.
PT   Nucleic acid encoding regulatory (p101) and catalytic (p120)
PT   subunits of a heterodimeric phosphatidylinositol 3' kinase - useful
PT   in treatment and diagnosis of immune system disorders, e.g.
PT   arthritis, cancer and Alzheimer's disease
PS   Example IX; Fig 1A-D; 75pp; English.
CC   This sequence encodes a novel adapter subunit (p101, also known as the
CC   regulatory subunit) from porcine phosphoinositide 3 hydroxykinase
CC   (PI3K) which is regulated by beta-gamma subunits of trimeric G-protein.
CC   Nucleic acid coding for p101 or p120, or their fragments, are used
CC   as probes and primers for identifying p101 or p120 gene mutations,
CC   allelic variations or regulatory defects, particularly for the diagnosis
CC   of activation disorders (or susceptibility) in cells of the
CC   haematopoietic system. The related proteins, antibodies, agonists and
CC   antagonists can be used similarly. The p101 and p120 proteins, peptides
CC   or fusion proteins are used to treat or screen for potential agents for
CC   treating immune disorders, particularly inflammation, e.g. arthritis,
CC   septic shock, adult respiratory distress syndrome, pneumonia, asthma,
CC   allergies, reperfusion injury, atherosclerosis, Alzheimer's disease and
CC   cancer.
SQ   Sequence   4692 BP;    916 A;    1559 C;    1394 G;    823 T;

  Query Match              3.8%;  Score 41.8;  DB 1;  Length 4692;
  Best Local Similarity   55.1%;  Pred. No. 0.33;
  Matches  102;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy     855 ccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttggaagacct 914
           ||| ||     || || || |  || | | | |  | ||  | |||||||| ||   |||
Db    1232 CCCCATCCCCGTCGCCAGGTGCTACACCTCACCTCGTGAACCAGGACAGCTTCGACATCCT 1291

Qy     915 ggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgcaatacgaaga 974
           | || || || | |  | ||| || |  | | ||          |||| |  |||| ||
Db    1292 GTAGGAAATCCTGCTCAAGGAGTAGGAGCTCCTCCAGCCAGAGATCCTGGACGACGAGGA 1351

Qy     975 ggaa-agactgaaggccaggagggagagcgcgaggccccagccggagtttgtgctgccca 1033
           |||  ||   || |   | |||| ||||  |  || | | | |||    || || |
Db    1352 GGACGAGGACGAGGAGGACGAGGAAGAGGACTTGGACGCCGACGGCCACTGCGCGGAGAG 1411

Qy    1034 ggtct 1038
           || ||
Db    1412 GGACT 1416


Search completed: February  8, 2000, 20:43:04
Job time: 832 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 07:12:44 ; Search time 5888.69 Seconds
(without alignments)
-2183.618 Million cell updates/sec

Title: US-09-174-937-4
Perfect score: 4235
Sequence: 1 gtcgacccacgcgtccgcgg.....aaaaaaaaaaggggcggccgc 4235

Scoring table: IDENTITY_NUC

Searched: 921193 seqs, -1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold : 1642396

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_htg5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4029.8	95.2	5654	9	AB002365	AB 2365 Human mRN
c 2	3226.4	76.2	169799	44	AC[illegible]	[illegible] Homo sapi
c 3	684	16.2	58386	42	AC[illegible]	[illegible] Homo sapi
4	286.2	6.8	1386	5	AF228453	AF 28453 Sequence
5	286.2	6.8	2382	11	U80[illegible]	U8[illegible] Human sapien
6	283	6.7	2213	12	AF[illegible]	AF[illegible] Mus muscu
7	206.2	4.9	58386	42	AC[illegible]	AC[illegible] Homo sapi
c 8	203.4	4.8	217	15	G[illegible]	G[illegible] Human STS S
9	202.4	4.8	169798	44	AC[illegible]	AC[illegible] Homo sapi
c 10	117	2.8	117	15	G[illegible]	G[illegible] Human STS S
c 11	97.2	2.3	114467	11	AF[illegible]	AF[illegible] Homo sapi
c 12	92.8	2.2	169868	11	AC[illegible]	AC[illegible] Homo sapi
c 13	92	2.2	179301	41	AC[illegible]	AC[illegible] Homo sapi
14	89.6	2.1	88215	40	AC[illegible]	AC[illegible] Homo sapi
c 15	85.6	2.0	78170	42	AC[illegible]	AC[illegible] Drosophil
16	81.2	1.9	34836	40	AC[illegible]	AC[illegible] Homo sapi
17	80.8	1.9	7218	5	166494	166494 Sequence 14
c 18	80.2	1.9	180221	40	AC[illegible]	AC[illegible] Homo sapi
c 19	79.6	1.9	156839	40	AC[illegible]	AC[illegible] Homo sapi
20	79.4	1.9	197294	42	AC[illegible]	AC[illegible] Homo sapi
21	79.4	1.9	100000	9	AL[illegible]	AL[illegible] Homo sapi
22	79.4	1.9	139229	25	AL[illegible]	AL[illegible] Homo sapi
23	79.4	1.9	120953	44	AC[illegible]	AC[illegible] Homo sapi
24	77.8	1.8	129756	11	HS[illegible]	AL[illegible] Human DNA
c 25	77.8	1.8	122460	41	AC[illegible]	AC[illegible] Homo sapi
26	77.6	1.8	152027	16	ON[illegible]	AL[illegible] BAC seque
27	77.6	1.8	198436	42	AC[illegible]	AC[illegible] Homo sapi
c 28	77.6	1.8	160969	44	AC[illegible]	AC[illegible] Homo sapi
29	77	1.8	169583	42	AC[illegible]	AC[illegible] Homo sapi
c 30	76.6	1.8	172093	42	AC[illegible]	AC[illegible] Homo sapi
c 31	76.2	1.8	160359	40	AC[illegible]	AC[illegible] Homo sapi
32	76.2	1.8	1754	40	AF[illegible]	AF[illegible] Homo sapi
33	75.6	1.8	53681	43	AC[illegible]	AC[illegible] Homo sapi
34	75	1.8	70519	41	AC[illegible]	AC[illegible] Homo sapi
c 35	74.6	1.8	200543	32	HSAS[illegible]	AL[illegible] Homo sapi
36	74.4	1.8	126150	11	AC[illegible]	AC[illegible] Homo sapi
37	74.4	1.8	140171	44	AC[illegible]	AC[illegible] Homo sapi
c 38	73.6	1.7	127735	11	HS[illegible]	AL[illegible] Human DNA
c 39	73.2	1.7	130817	40	AC[illegible]	AC[illegible] Homo sapi
40	73	1.7	180588	44	AC[illegible]	AC[illegible] Homo sapi
41	73	1.7	167819	44	AC[illegible]	AC[illegible] Homo sapi
42	72.8	1.7	169316	39	ON[illegible]	AL[illegible] BAC seque
43	72.6	1.7	173627	40	AC[illegible]	AC[illegible] Homo sapi
44	72.4	1.7	143427	11	AC[illegible]	AC[illegible] Human BAC
45	72.4	1.7	56959	11	HS[illegible]	AL[illegible] Human DNA

ALIGNMENTS

RESULT 1
AB002365
LOCUS AB002365 5654 bp mRNA PRI 09-FEB-1999
DEFINITION Human mRNA for KIAA[illegible] gene, partial cds.
ACCESSION AB002365
VERSION AB002365.1 GI:2224674
KEYWORDS KIAA0367.
SOURCE Homo sapiens male brain cDNA to mRNA, clone:pBluescriptII SK plus clone:HH0141.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp; Tel:+81-438-52-3930,
            Fax:+81-438-52-3931;
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescript II SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA0367"
     CDS             <1..2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20820.1"
                     /db_xref="GI:2224875"
                     /translation="ALFTGDR-LSIENFALVPDALLASDTCLDISEAAFDHSFSDASG
                     LNTSTGTIDDMSKLTLSEGHPETPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDLL
                     REPEHFLYGGDPPLEFDSIKQSIAPYTPPFDLSYLTEPAQSAETIEEAGSPEDESLGC
                     RAAEIVLSALPDRRSEGNQAEIKNRLPGSQLAVLHIREDPESVYLPVGAGSNILSPSN
                     VDWEVFTDNSDLPAGGDTGPPNGASKEIPELEEEKTIPTKEPEQIKSEYKEEPCTEKN
                     EDRHALHMDYILVNREENSHSKPETCEERESIAELELYVGSKETGLQGTQLASFPDTC
                     QPASLNERKGLSAEKMSSKGCGSSFESPAQDQSWMFLGHSEVGDPSLDARDSGPGWS
                     GKTVEPFSELGLGEGPQLQILEEMKPLESLALEEASGPVSQSQKSKSPGRAGPDAVTH
                     DSEWEMLSPQPVQKNMIRDTPMEEETEFLEIGTPISPPNGLLSEDVGMEIPFEEPVLS
                     PSAADMRPEPPNSLDLNDTHPRRIKLTAPNINLSLDQSEGSILSDDNLDSPDEIDINV
                     DFLDTPDFADSFEYIGHDPIAMKDSGQESDSIPPYTAEEEFEDNPLWPTVVIQEQEP
                     IDMKVIEPYRRVISHGGYYGICLNAIIVFAACFLPDSSRADYHYVMENLFLYVISTLE
                     LMVAEDYMIVYLNGATPKKEDAPAPLDEEMIPDD"
BASE COUNT     1564 a   1188 c   1198 g   1604 t
ORIGIN


  Query Match             95.2%;  Score 4029.8;  DB 9;  Length 5654;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 4158;  Conservative    0;  Mismatches    2;  Indels   97;  Gaps    3;

Qy      16 cacggaaaatggaggaggagacagagttccttgagctcggaaccaggatatcaagaccaaa 75
           |||||||||||||||||||||||||||||| ||  ||||||||||||||||||||||||||
Db    1407 CACGGAAATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGAACCAGGATATCAAGACCAAA 1466

Qy      76 tggactactgtcagaggatgtaggaatggacatccccttgaagagggcgtgctgagtcc 135
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1467 TGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTGAAGAGGGCGTGCTGAGTCC 1526

Qy     136 cagtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcc 195
           |||||||||||||||||||||||||||||| ||  ||||||||||||||||||||||||||
Db    1527 CAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCC 1586

Qy     196 tggaaaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatc 255
           |||||||||||||||||||||||||||||| ||  ||||||||||||||||||||||||||
Db    1587 TGGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATC 1646

Qy     256 tattctctctgatgataacttggacagtccaatgaaattgacatcaatgtggatgaact 315
           |||||||||||||||||||||||||||||| ||  ||||||||||||||||||||||||||
Db    1647 TATTCTCTCTGATGATAACTTGGACAGTCCAATGAAATTGACATCAATGTGGATGAACT 1706

Qy     316 tgatacccccgatgaagcagattcttttgagtacactgaccatgatcccacaccaacaccaa 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1707 TGATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGACCATGATCCCACACCAACACCAA 1766

Qy     376 agattctggccaagagtcagagtt------atatatatttttttttttttttttttttt 435
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1767 AGATTCTGGCCAAGAGTCAGAGTCTATTCCCATATATATTTTTTTTTTTTTTTTTTTTTT 1826

Qy     436 caaccggctttggatgacagtggtccattggagaacaagagcagccacatgacacaaggtt 495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1827 CAACCGGCTTTGGAGGACAGTGGTCCATTGGAGAACAAGAGCAGCCACATGACACAAGGTT 1886

Qy     496 catcgagccctacaggagagtcatttctcacggaggagattcaggatactatttaaacct 555
           |||||||||||||||||||||||||||| |||||||          ||||||||||||||
Db    1887 CATCGAGCCCTACAGGAGAGTCATTTCTCACGG---------AGGATACTATTTTAAACCT 1937

Qy     556 tctaaatgccatcattgtgtttgccgcctgtttttctgccagacagcagtctcaaaaaatta 615
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1938 TCTAAATGCCATCATTGTGTTTGCCGCCTGTTTTTCTGCCAGACAGCAGTCTCAAAAATTA 1997

Qy     616 ccactatgtcatggaaaatcttttcctatatgtaataagtactttagagttgatgtgtagt 675
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |||
Db    1998 CCACTATGTCATGGAAAATCTTTTCCTATATGTAATAAGTACTTTAGAGTTTATGTGTAGT 2057

Qy     676 tgaagactatatgattgtgtacttgaatggtgcaacccccaagaaaaaaggagtatcaaaat 734
           |||||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||
Db    2058 TGAAGACTATATGATTGTGTACTTGAATGGTGAAGCCCCAAGAAAAAAGGAGTATCAAAAT 2117

Qy     735 taggctggatgaagaaatgctaccagatgattgacagacggttgaaaaagtgttttaaaat 794
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2118 TAGGCTGGATGAAGAAATGCTACCAGATGATTGACAGACGGTTGAAAAAGTGTTTTAAAAT 2177

Qy     795 cattcatcattgttcatccatcttgttgtcatgaagaaatgcttgctgtgaaaaagcctt 854
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2178 CATTCATCATTGTTCATCCATCTTGTTGTCATGAGAACAATGCTTGCTGTGAAAAAGCCTT 2237

Qy     855 ttataagttcaaaattcagcagtaaaattaaatatgtcaatagcttatgaaaattaaaat 914
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2238 TTATAAGTTCAAAATTCAGCAGTAAAATTAAATATGTCAATAGCTTATGAAAATTAAAAT 2297

Qy     915 ggctgatccccaatggattgcatgtg attccagagaggcatcatga---------------- 960
           |||||||||||||||||||||||||| |||||||||||||||||||                
Db    2298 GGCTGATCCCAATGGATTGCATCTATTCCAGAGAGGCATCATAAA---------------- 2357

Qy     960 ---------------------------------------------------------- 960

Db    2358 TGAGGGAAGCATCAGAGGCAGCTAAAACTAGCTGCCTTTACAATGATCCTGAAAATTTTT 2417

Qy     960 ------------atattgacttgaaggctgaaagaaaaagccttagttgccatgtgaaaa 1007
                       |||||||||||||||||||||||||||||||||||||||||||||||
Db    2418 CTATGGAGAAGGATATTGACTTGAAGCTGAAAGAAAAGCCTTAGTTGCCATGTGAAAA 2477

Qy    1008 aagaggatgcttttctggttcatggttctgttgaaacatatctacctgaaaatgatgaaa 1067
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Db    2478 AAGAGGATGCTTTTCTGGTTCATGGTTCTGTTGAAACATATCTACCTGAAAATGATGAAA 2537

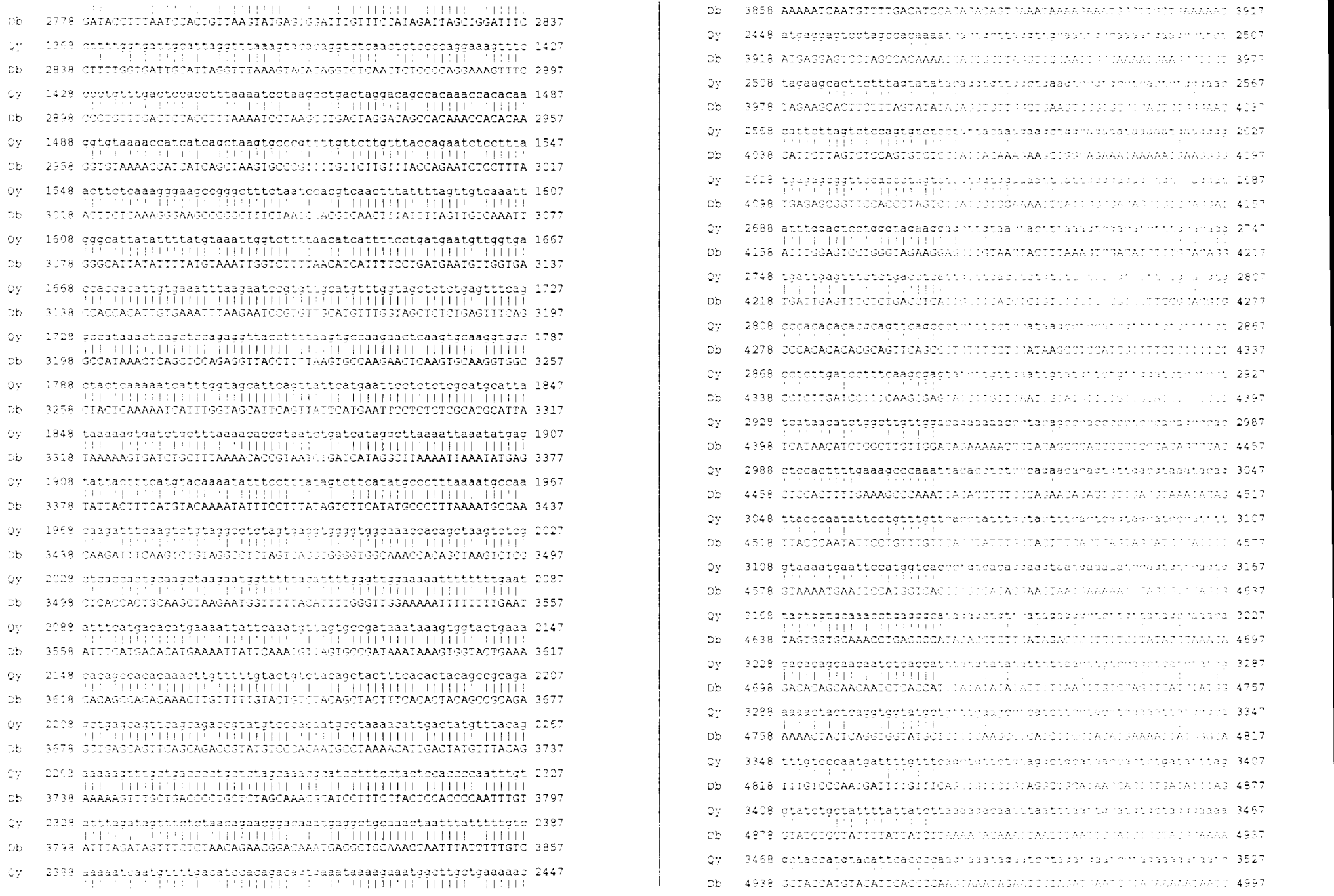
Qy    1068 ctgatgttacctttttccactttgcactacctggtgccattctaaatttctaaaaaaaaa 1127
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    2538 CTGATGTTACCTTTTTCCACTTTGCACTACCTGGTGCCATTCTAAATTTCTAAATTTTTAA 2597

Qy    1128 aatagaaagtttgtttactcttaagatattttatgaaatttgtgtgtactttcctatttta 1187
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2598 AATAGAAAGTTTGTTTACTCTTAAGATATTTTATGAAATTTGTGTGTACTTTCCTATTTTA 2657

Qy    1188 ccaattatgtgcctcaaagatttttagttgagccttagcaagaaaagtaggagcttttgaaa 1247
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2658 CCAATTATGTGCCTCAAAGATTTTAGTTGAGCCTTAGCAAGAAAAGTAGGAGCTTTTGAAA 2717

Qy    1248 caatacttcattaacacggtgtagtgatactttgtcccttagactggtgtttatgcagtaa 1307
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2718 CAATACTTCATTAACACGGTGTAGTGATACTTTGTCCCTTAGACTGGTGTTTATGCAGTAA 2777

Qy    1308 gatacctttaatccactgttaagtatggtggatttgtttccatagattaatgtgattttt 1367
```



```
Qy    3529 ttttgtttgataaatgatgatggaggtaaaatttacatgatttagctctctagctcttgca 3588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4998 CCTAAGCAGATAGGTAGACAGAGGIAAACAI.TACATGATTTAGCTCTCTAGCTCTTGCA 5057

Qy    3589 ctctgaaacattcttgctttggttctgaattctaggaactgctttgcatttctcctataga 3647
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5058 CTCTGAACATTCTTGCTTTGGTTCTGACTTCTAGGAACTGCTTTGCATTTCTCCTATAGA 5117

Qy    3648 tattgtattaaaaaaatcaaaggtctattggtcaaaaagcattgttttctcaaagctcctt 3707
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5118 TTTGTAGTTAAGGGAACCAAGGGGTCATTGG.CAAAAGCATTGTTTCTCAAAGCTCCTT 5177

Qy    3708 gattaagagaaagaaacagaaatttgcacagaaatagtgtcaaggagtgagaaagtttgt 3767
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5178 GATTAACACAAACAACACAAATTTCTACACAACATAGTTCTCAAGGAGTGAGAAAGTTTGT 5237

Qy    3768 ttgagggcagtagctcagtgtggaagaaaatcctgaagtttctgttgaagccatacaatg 3827
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5238 TTGAGGGCAGTAGCTCAGTGTGGAAGAAAATCTTGAAGTTTCTGTTGAAGCCATACAATG 5297

Qy    3828 ttctatgggttactctctaagacattctctgaggtgtgtgaggaagtcactactcctag 3887
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5298 TTCTATGGGTTACTCTCTTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAG 5357

Qy    3888 cctttgttaagatgtaatttttaaatattcagttatggtactatgtttgcaactctcgtct 3947
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5358 CCTTTGTTAAGATGTAATTTTAAATATTCAG.TATGGTACTATGTTTGCAACTCTCGTCT 5417

Qy    3948 tatcacaatgcctcagtagtttgttccctaaaaacatttagatgtgcacaaattaatct 4007
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5418 TATCACAATGCCTCAGTAGTTTGTTCCCTTAGAAACATTTACATGTGCACAAATTAATCT 5477

Qy    4008 tttatatatctaaaggtttttctatcatgcattggattgctcagaataaagtgtctgtta 4067
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5478 TTTATATATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTA 5537

Qy    4068 gacttcgttttggtaaaataaattctccataatatagattaataatataaaagtctttaat 4127
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5538 GACTTCGTTTTGGTAAATAAATTCTCCATAATATAGATTAATAATATAAAAGTCTTTAAT 5597

Qy    4128 gacacaatatatctatatagcctcactgtataattcagaaaataaaaattgattctgc 4184
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5598 GACACAATATATCTATATAGCCTCACTGTATAATTCAGAAATAAAAATTGATTCTGC 5654


RESULT    2
AC009556/c
LOCUS       AC009556    169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dupa,ie,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira P., FitzHugh,W., Forrest,C.,
            Fulton,R., Gaje,C., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas A., Lehoczky J., Liou C., Locke K., Macdonald P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Mulla,M., Morris,G., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Perlin,M.,
            Peterson,K., Pollara,V., Riley,P., Roberts,D., Roy,A., Berry,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:4114934.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *              gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *              gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *              gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *              gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *              gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *              gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *              gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *              gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *              gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *              gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *              gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *              gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *              gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *              gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *              gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *              gap of unknown length
            *    58264    65216: contig of 6953 bp in length
```

```
            *                    gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                    gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                    gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                    gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                    gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                    gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                    gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                    gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                    gap of unknown length
            *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN

  Query Match             76.2%;  Score 3226.4;  DB 44;  Length 169798;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3227;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     958 caatattgacttgaaagctgaaagaaaagccttagttggccatgctggaaagaaggaggatgc 1017
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70225 CAGTATTGACTTGAAGCTGAAAGAAAAGCTTTAGTTGGCCATGCTGGAAGAAGAGGATGC 70166

Qy    1018 ttttctggttcatggtttctgttgaaacatatttacctgaaagagacagggctgatgttac 1077
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70165 TTTTCTGGTTCATGGTTCTGTTGAAACATATTTACCTGAAAGAGACAGGGCTGATGTTAC 70106

Qy    1078 ctttttccactttgcactacctggtgccattctaaatttctaaggggaaaaatagaaagt 1137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70105 CTTTTTCCACTTTGCACTACCTGGTGCCATTCTAAATTTCTAAGGGGAAAAATAGAAAGT 70046

Qy    1138 ttgtttactcttaagatatttttatgaaattgtgtgtactttcctatttgccaattatgt 1197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70045 TTGTTTACTCTTAAGATATTTTATGAAATTGTGTGTACTTTCCTATTTGCCAATTATGT 69986

Qy    1198 gcctcaaagattttagttgagccttagcaagaaagtaggaccttccatttcaatacttca 1257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69985 GCCTCAAAGATTTTAGTTGAGCCTTAGCAAGAAAGTAGGACCTTCCATTTCAATACTTCA 69926

Qy    1258 ttaacacggtgtagtgatactttgtccctaaactggtgtttaccagtaagatacctttta 1317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69925 TTAACACGGTGTAGTGATACTTTGTCCCTTAGACTGGTGTTTACCAGTAAGATACCTTTA 69866

Qy    1318 atccactgttaagtatgagtggatttgttttctatagattagctggatttccttttggtga 1377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69865 ATCCACTGTTAAGTATGAGTGGATTTGTTTTCTATAGATTAGCTGGATTTCCTTTTGGTGA 69806

Qy    1378 ttgcattaggtttaaagtacacaggtctcaactctccccaggaaagtttccccctgtttga 1437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69805 TTGCATTAGGTTTAAAGTACACAGGTCTCAACTCTCCCCAGGAAAGTTTCCCCTGTTTGA 69746

Qy    1438 ctccacctttaaaatcctaagcctgactaggacagccacaaaccacacaaggtgtaaaac 1497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69745 CTCCACCTTTAAAATCCTAAGCCTGACTAGGACAGCCACAAACCACACAAGGTGTAAAAC 69686

Qy    1498 catcatcagctaagtgcccgttttgttcttgtttaccagaatctcctttaacttctcaaa 1557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69685 CATCATCAGCTAAGTGCCCGTTTTGTTCTTGTTTACCAGAATCTCCTTTAACTTCTCAAA 69626

Qy    1558 ggaaagcgggctttctaatccaagtcgaattttatttaatttcgaattttcgaattata 1617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69625 GGAAAGCGGGCTTTCTAATCCAAGTCGAATTTTATTTAATTTCGAATTTTCGAATTATA 69566

Qy    1618 ttttatgtaaaattggtcttttaa atcatttt tctaatgaat.tttgttaa'tacta'att 1677
           |||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||||
Db   69565 TTTTATGTAAATTGGTCTTTTAATCATCATTTTTCTAATGAATTTTGTTAATACTAATT 69506

Qy    1678 gtgaaatttaagaatccgtgttg ctgtttgtagtgt tctttgtttt aaaattaaaat 1737
           ||||||||||||||||||||||| |||||||||||| |||||||||| |||||||||||||||||||
Db   69505 GTGAAATTTAAGAATCCGTGTTGTTTGTTTTTGTGTACTTTTTTGTTTTAAATTAAAAT 69446

Qy    1738 cagctccagagatttaaatttttgctgtcctggctctaatgtatttgtttaaaa 1797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69445 CAGCTCCACACGTTACCTTTTATTTTTTATTTTTTTTTTTTTTTTTTTTTTAAAA 69386

Qy    1798 atcatttggtagcattcagttattcatgaatttctttcgaatgcattatttaaaaattga 1857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69385 ATCATTTGGTAGCATTCAGTTATTCATGAATTTCTTTCGAATGCATTATTAAAAATTGA 69326

Qy    1858 tctgctttaaaaacaccgtaatctgatcatggcttaaaattaaatatgagtattaattttc 1917
           ||||||||||| ||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Db   69325 TCTGCTTTAAAACACCGTAATCTGATCATGGCTTTAAATTAAATATGAGTATTAATTTTC 69266

Qy    1918 atgtagcaaaaatatttcctttatgtgtctttatgtgctcttaaaagtgcaagcaaaaattttga 1977
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69265 ATGTAGCAAAAATATTTCCTTTATGTGTCTTTATGTGCTCTTAAAAGTGCAAGCAAAAATTTTGA 69206

Qy    1978 agtctctaggcctctagtgagggtaaaatgtcctaatgaaatgaatgtgtgtgtgtgtgtgtgt 2037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69205 AGTCTCTAGGCCTCTAGTGAGGTTAGGTCTAAACTAATGAATTAATTTGTGTGTGTGTGTGT 69146

Qy    2038 caagctaagaatggttttacatttgtttctaaaaatttttttttgtatttttatgtgtgt 2097
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69145 CAAGCTAAGAATGGTTTTTACATTTGTTTCTAAAAATTTTTTTTGTAATATTTTATGTGT 69086

Qy    2098 acatgaaaattattcaaatgttaatgcctatataataatgatgtatttatgaatatgtgtgt 2157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69085 ACATGAAAATTATTCAAATGTTAATGCCTATATAATAATGATGTATTTATGAATATGTGTGT 69026

Qy    2158 caaacttgttttttgtactgtctgcagctgtttcagaatgaagatgtgtgtgtgtgtgtgt 2217
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69025 CAAACTTGTTTTTGTACTGTCTACAGCTGTTTCAGAATGAAGATGTGTGTGTGTGTGTGT 68966

Qy    2218 tcagcagacggtatgtcccacctttctaaatattgattcttatttttttttgtgtgtgtgt 2277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68965 TCAGCAGACGGTATGTCCCACCTTTCTAAATATTGATTCTTATTTTTTTTGTGTGTGTGTGT 68906

Qy    2278 ctgacccctgctctagcaaacgtatcctttcctactctcccctcatttttattatgtgt 2337
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68905 CTGACCCCTGCTCTAGCAAACGTATCCTTTCCTACTCTCCCCTCATTTTTATTATGTGT 68846

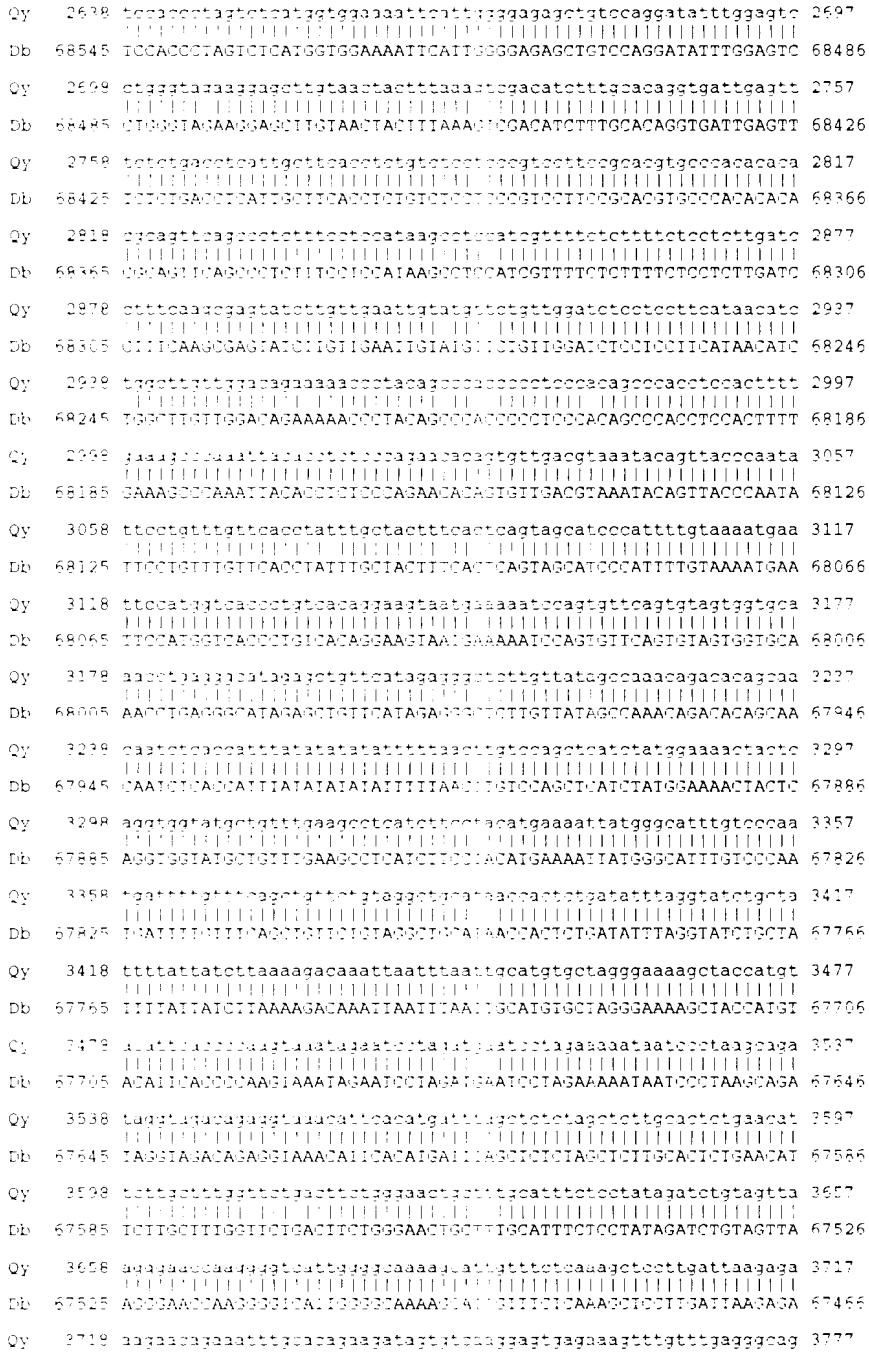
Qy    2338 tttctctaacagaacggacaaattgagcttcaaactaatttattttttctgaaaaattgaat 2397
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68845 TTTCTCTAACAGAACGGACAAATTGAGCTTCAAACTAATTTATTTTTTCTGAAAAATTGAAT 68786

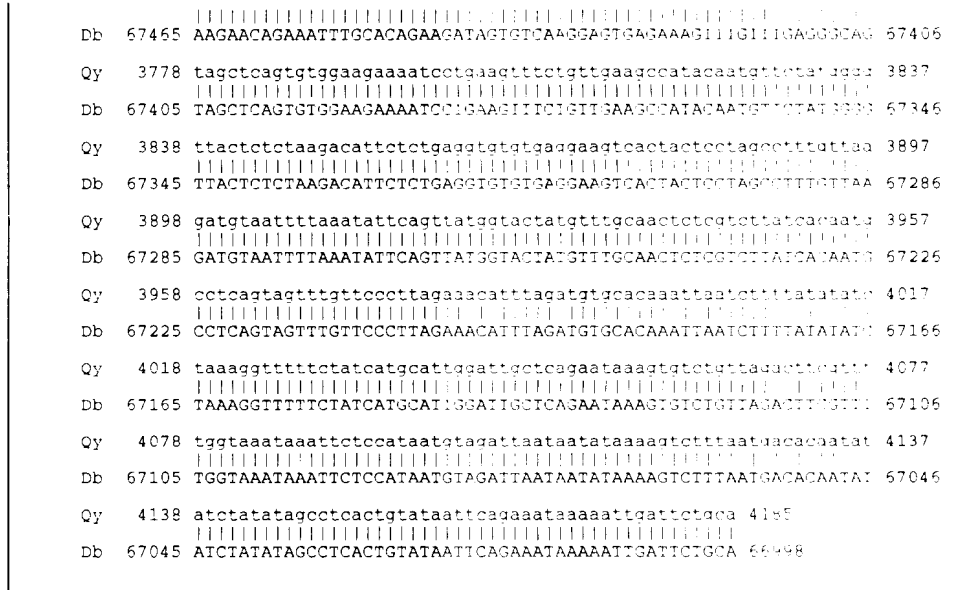
Qy    2398 gttttgacatccacagacagtgaaatatgtgaaatgtgtttgtttgaaaaatgtttttttct 2457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68785 GTTTTGACATCCACAGACAGTGAAATATGTGAAATGTGTTTGTTTGAAAAATGTTTTTTTCT 68726

Qy    2458 ctagccacaaaatcactgttagtttgcaatttccaaaatgaaattcctttttaaaaaaaaat 2517
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68725 CTAGCCACAAAATCACTGTTAGTTTGCAATTTCCAAAATGAAATTCCTTTTTAAAAAAAAT 68666

Qy    2518 tctttagtatatacagagtattgg tgaaatgcctg gtgatttgaaaaatttttttttt 2577
           |||||||||||||||||||||||| ||||||||||| |||||||||||||||||||||||||||||
Db   68665 TCTTTAGTATATACAGGGTGTTGGTTGAAATGCCTGTGCTTGAATTTTTGAAAATTTTTTTT 68606

Qy    2578 ctccagtgtctcctattacaaagaaatgtgtgaaatgaaaaatgaaaaatgtgtgaaatgt 2637
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68605 CTCCAGTGTCTCCTATTACAAAGAAATGTGTGAAATGAAAAATGAAAAATGTGTGAAATGT 68546
```





```
RESULT   3
AC012249/c
LOCUS       AC012249    58386 bp    DNA             HTG       21-OCT-1999
DEFINITION  Homo sapiens clone 16_N_23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC012249
VERSION     AC012249.1  GI:6091795
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone 16_N_23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Berkeley,E., Bessalaieff,T., Chandler,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Cullmore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,E., Jones,C., Kann,L., Kapatos,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Moldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,S.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,C.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,T., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker, Smit, A. F. A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This record contains 79 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
```

```
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      799: contig of 799 bp in length
*                    gap of unknown length
*      800     1536: contig of 737 bp in length
*                    gap of unknown length
*     1537     2265: contig of 729 bp in length
*                    gap of unknown length
*     2266     2972: contig of 707 bp in length
*                    gap of unknown length
*     2973     3725: contig of 753 bp in length
*                    gap of unknown length
*     3726     4475: contig of 750 bp in length
*                    gap of unknown length
*     4476     5247: contig of 772 bp in length
*                    gap of unknown length
*     5248     5992: contig of 745 bp in length
*                    gap of unknown length
*     5993     6730: contig of 738 bp in length
*                    gap of unknown length
*     6731     7474: contig of 744 bp in length
*                    gap of unknown length
*     7475     8210: contig of 736 bp in length
*                    gap of unknown length
*     8211     8928: contig of 718 bp in length
*                    gap of unknown length
*     8929     9640: contig of 712 bp in length
*                    gap of unknown length
*     9641    10387: contig of 747 bp in length
*                    gap of unknown length
*    10388    11120: contig of 733 bp in length
*                    gap of unknown length
*    11121    11879: contig of 759 bp in length
*                    gap of unknown length
*    11880    12648: contig of 769 bp in length
*                    gap of unknown length
*    12649    13441: contig of 793 bp in length
*                    gap of unknown length
*    13442    14200: contig of 759 bp in length
*                    gap of unknown length
*    14201    14936: contig of 736 bp in length
*                    gap of unknown length
*    14937    15720: contig of 784 bp in length
*                    gap of unknown length
*    15721    16404: contig of 684 bp in length
*                    gap of unknown length
*    16405    17121: contig of 717 bp in length
*                    gap of unknown length
*    17122    17847: contig of 726 bp in length
*                    gap of unknown length
*    17848    18601: contig of 754 bp in length
*                    gap of unknown length
*    18602    19316: contig of 715 bp in length
*                    gap of unknown length
*    19317    20058: contig of 742 bp in length
*                    gap of unknown length
*    20059    20787: contig of 729 bp in length
*                    gap of unknown length
*    20788    21518: contig of 731 bp in length
*                    gap of unknown length
*    21519    22254: contig of 736 bp in length
*                    gap of unknown length
*    22255    23015: contig of 761 bp in length
*                    gap of unknown length
*    23016    23737: contig of 722 bp in length
*                    gap of unknown length
*    23738    24453: contig of 716 bp in length
*                    gap of unknown length
*    24454    25192: contig of 739 bp in length
*                    gap of unknown length
*    25193    25937: contig of 745 bp in length
*                    gap of unknown length
*    25938    26683: contig of 746 bp in length
*                    gap of unknown length
*    26684    27431: contig of 748 bp in length
*                    gap of unknown length
*    27432    28179: contig of 748 bp in length
*                    gap of unknown length
*    28180    28941: contig of 762 bp in length
*                    gap of unknown length
*    28942    29667: contig of 726 bp in length
*                    gap of unknown length
*    29668    30386: contig of 719 bp in length
*                    gap of unknown length
*    30387    31105: contig of 719 bp in length
*                    gap of unknown length
*    31106    31889: contig of 784 bp in length
*                    gap of unknown length
*    31890    32643: contig of 754 bp in length
*                    gap of unknown length
*    32644    33422: contig of 779 bp in length
*                    gap of unknown length
*    33423    34169: contig of 747 bp in length
*                    gap of unknown length
*    34170    34901: contig of 732 bp in length
*                    gap of unknown length
*    34902    35643: contig of 742 bp in length
*                    gap of unknown length
*    35644    36403: contig of 760 bp in length
*                    gap of unknown length
*    36404    37119: contig of 716 bp in length
*                    gap of unknown length
*    37120    37831: contig of 712 bp in length
*                    gap of unknown length
*    37832    38555: contig of 724 bp in length
*                    gap of unknown length
*    38556    39332: contig of 777 bp in length
*                    gap of unknown length
*    39333    40130: contig of 798 bp in length
*                    gap of unknown length
*    40131    40891: contig of 761 bp in length
*                    gap of unknown length
*    40892    41629: contig of 738 bp in length
*                    gap of unknown length
*    41630    42372: contig of 743 bp in length
*                    gap of unknown length
*    42373    43156: contig of 784 bp in length
*                    gap of unknown length
*    43157    43889: contig of 733 bp in length
*                    gap of unknown length
*    43890    44610: contig of 721 bp in length
*                    gap of unknown length
*    44611    45411: contig of 801 bp in length
*                    gap of unknown length
*    45412    46177: contig of 766 bp in length
*                    gap of unknown length
*    46178    46940: contig of 763 bp in length
*                    gap of unknown length
*    46941    47711: contig of 771 bp in length
*                    gap of unknown length
*    47712    48462: contig of 751 bp in length
*                    gap of unknown length
*    48463    49258: contig of 796 bp in length
*                    gap of unknown length
*    49259    50004: contig of 746 bp in length
*                    gap of unknown length
*    50005    50781: contig of 777 bp in length
*                    gap of unknown length
*    50782    51487: contig of 706 bp in length
*                    gap of unknown length
*    51488    52199: contig of 712 bp in length
*                    gap of unknown length
```

```
            *    52200    52928: contig of 729 bp in length
            *           gap of unknown length
            *    52929    53736: contig of 808 bp in length
            *           gap of unknown length
            *    53737    54547: contig of 811 bp in length
            *           gap of unknown length
            *    54548    55305: contig of 758 bp in length
            *           gap of unknown length
            *    55306    56070: contig of 765 bp in length
            *           gap of unknown length
            *    56071    56822: contig of 752 bp in length
            *           gap of unknown length
            *    56822    57607: contig of 785 bp in length
            *           gap of unknown length
            *    57608    58386: contig of 779 bp in length.

  Query Match             16.2%;  Score 684;  DB 42;  Length 58386;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-121;
  Matches  687;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy   2984 ccacctccactttgaaaggcccaaattacacctctcccagaacacagtgttgacgtaaat 3043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29663 CCCACTCCACTTTGAAAGCCCAAATTACACCTCTCCCAGAACACAGTGTTGACGTAAAT 29604

Qy   3044 acagttacccaatattctgtttgttcactatttgctacttttcactcagtagcatccca 3103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29603 ACAGTTACCCAATATTCCTGTTTGTTCACTTTATTGCTACTTTTCACTCAGTAGCATCCCA 29544

Qy   3104 ttttgtaaaatgaattccatggtcaccctgtcacaggaagtaatgaaaaatccagtgttc 3163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29543 TTTTGTAAAATGAATTCCATGGTCACCCTGTCACAGGAAGTAATGAAAAATCCAGTGTTC 29484

Qy   3164 actgtaatggtacaaacctgagggcataagctgttcatagagggctcttgttatagcca 3223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29483 ACTGTACTGGTGCAAACCTCAGGGCATAGACTGTTCATAGAGGGCTCTTGTTATAGCCA 29424

Qy   3224 aacagacacagcaacaatctcaccatttatatatatttttaacttgtccagctcatct 3283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29423 AACAGACACAGCAACAATCTCACCATTTATATATATTTTTAACTTGTCCAGCTCATCT 29364

Qy   3284 atggaaaactactcaggtggtatgctgtttgaagcctcatcttcctacatgaaaattatg 3343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29363 ATGGAAAACTACTCAGGTGGTATGCTGTTTGAAGCCTCATCTTCCTACATGAAAATTATG 29304

Qy   3344 agcatttgtcccaatgatttttgttcagctgttctgtaggctgcataaccactctgatat 3403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29303 GGCATTTGTCCCAATGATTTTTGTTCAGTTGTTCTGTAGGCTGCATAACCACTCTGATAT 29244

Qy   3404 ttaggtatctgctattttattatcttaaaagacaaattaatttaattgcatgtgctaggg 3463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29243 TTACCTATCTGCTATTTTATTATCTTAAAAGACAAATTAATTTAATTGCATGTGCTAGGG 29184

Qy   3464 aaaagctaccatgtacattcaccccaagtaaatagaatcctagatgaatcctagaaaaat 3523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29183 AAAAGCTACCATGTACATTCACCCCAAGTAAATAGAATCCTAGATGAATCCTAGAAAAAT 29124

Qy   3524 aatcctaagcagatagtagtagacagaggtaaaaattcacatgatttagctctctagctct 3583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29123 AATCCTAAGCAGATAGTAGTAGACAGAGGTAAAATTCACATGATTTAGCTCTCTAGCTCT 29064

Qy   3584 tgcactctgaacattcttgctttggttctgactctgggaactgctttgcatttctccta 3643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29063 TGCACTCTGAACATTCTTGCTTTGGTTCTGACTCTGGGAACTGCTTTGCATTTCTCCTA 29004

Qy   3644 tagatctgtagttaagggaaccaagggtca 3676
          |||||||||||||||||||||||||||||
Db  29003 TAGATCTGTAGTTAAGGGAACCAAGGGTCA 28971


RESULT   4
AR028453
LOCUS       AR028453     1396 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 22 from patent US 5858678.
ACCESSION   AR028453
VERSION     AR028453.1  GI:5940426
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G.
  TITLE     Apoptosis-regulating proteins
  JOURNAL   Patent: US 5858678-A 22 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1..1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    336 g    373 t
ORIGIN


  Query Match             6.8%;  Score 286.2;  DB 5;  Length 1386;
  Best Local Similarity   61.7%;  Pred. No. 2.5e-45;
  Matches  513;  Conservative    0;  Mismatches  303;  Indels   15;  Gaps    3;

Qy    199 gagaatcaagctcacagccccaaaatatcaatctttctctggaccaaaagtgaaaaatatat 258
          |||||  || || |  || ||| | | |  | |  | || | |  |  |  |  |  |
Db    360 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGCTCTGT 419

Qy    259 tctctctgatgataacttggacagtccagatgaaattgacatcaatatgtggatgaacttta 318
          | || ||||||      || |||    |  || |||||| |  |||     :
Db    420 ATTGTCAGATGATTTGGATGAAAGT---GGGGAGATTGACTTAGATGGCTTAGA   GAT 473

Qy    319 taccccctgatgaagcagattcttttgagtacactggccatgatcccacagccaacaaaaa 378
          |  | ||  |    ||   |||||||    |  || ||  | | | |
Db    474 ACCGTCAGAGAATAGTAATGAGTTTGAGTGGGAAGATGATCTTCCAAAACCTAAGACTAT 533

Qy    379 ttctggccaagagtcagagtctattccagaatatacggccgaaggaggaacaaaaggaaaa 438
          |  |       |  ||  || ||  |||| || | | | | |
Db    534 TGAAGTAATTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGAAGATTT 593

Qy    439 ccggctttggatgacagtggtcattgaggaacgagagcagcgcattgacatgaaggtcat 498
          || |  ||  |   |  ||||||||||| || |  | | | | | | | | | |
Db    594 ACGACGCTGGCGTATGTTCAGGATTGGAGAACAGGACCACAGGGTTGATATGAAGGTAAT 653

Qy    499 cgagccctacaggagagtcatttctcacggaggagattcaggatactatgaaaacgatct 558
          || |||| |  | ||| ||    || ||          |||| |||||||| | | |
Db    654 TGAACCCTATAAAAAAGTTATCAGCCATGG---------GGGATATTATGGGGATGGATT 704

Qy    559 aaatgccatcattgtgtttgccgcctgtttttctgccagacagcagtcgggcgaattaaaa 618
          |||||||||  |||| |||||  |||||  ||| |  | | | | | | | | | | |
Db    705 AAATGCCATTGTTGTATTTGCTGTCTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAT 764

Qy    619 ctatgtcatggaaaatcttttcctatatgtaataagtactttagagttgatagtagcttaa 678
          ||  | ||||  |||||||  |  | | | | | | | | | | | | | | | |
Db    765 ATACCTGATGGACAATCTTTTTAAATATGTTATTGGCACTTTGGAGCTATTAGTAGTATA 824

Qy    679 agactatatgattgtgtacttgaatgatgcagcccaagaaggaggataccaattgtatt 738
          | |||| |||||  || || ||  || ||||||||  |  |||| |  |||||  | |
Db    825 AAACTACATGATAGTTTATTTAAATGCTGCAACAACTCGAAGAAAAATGCCCAATTTTTT 884

Qy    739 ctggatgaagaaatgctaccagatgattgacagacggttgaggaaggaatttaaaatgat 798
          |||  | |  |||||  ||  |||   ||||| ||| |||| ||||  |  |  |
Db    885 ATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGAAAAATCTAAAATTTTT 944

Qy    799 catcattgttcatccatcttggttcatcagaacaatccttgctgtggacaccttttttat 858
          ||||||||  |||||  ||||||  | | | | |  |  |  |  |  |  |  |
Db    945 AATCATTGTACATCCTTCTTGGTTTATCAGAACACTTCTGGCTGTTACAATATTATTTAT 1004

Qy    859 aagttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaaactttttttt 918
          || || |||||||||   |||||| || ||  ||  | | | | | | | | | | |
Db   1005 TAGCTCGAAATTCAGCCAAAAAATTACATACGTGTTTAATTTGGCAAAACTACAATAA  1064

Qy    919 gatcccaatggattgcatccacattccagagagcatcatcaatattgactttaaagcttat 978
          ||||  |||| | | |  ||| |||||  |||| |  |  ||||  |  |  |  |
Db   1065 TGTCCCCATGGAATACGTTGGCATACCAGAATCCATAAAACAACTTTATCCAAATATTAT 1124
```

```
Qy    979 agaaaaagccttagttggccatgctggaagaagaggatgcttttctggttca 1029
          |||| |   |    | ||     |||  |  |   ||| || ||
Db   1123 TGGAAAACAAGATGAACCGAAAAATGAACAGTAAGTTTGGCATCTAGTCCA 1175


RESULT   5
HSU15173
LOCUS       HSU15173     2382 bp    mRNA            PRI       02-FEB-1998
DEFINITION  Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2)
            mRNA, complete cds.
ACCESSION   U15173
VERSION     U15173.1  GI:558843
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 212 to 1156)
  AUTHORS   Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K.,
            Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.
  TITLE     Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
            of cellular proteins
  JOURNAL   Cell 79, 341-351 (1994)
  MEDLINE   95042730
REFERENCE   2  (bases 1 to 2382)
  AUTHORS   Boyd,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1994) Janice M. Boyd, Institute for Molecular
            Virology, St. Louis University Health Sciences Center, 3681 Park
            Avenue, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B-cell, S.Elledge, B-cell, J.Ambrus; fetal
                     brain, M.Green"
                     /cell_type="B cells; brain"
                     /dev_stage="fetus (brain)"
     gene            1. .2382
                     /gene="BNIP2"
     CDS             212. .1156
                     /gene="BNIP2"
                     /codon_start=1
                     /product="BCL2/adenovirus E1B 19kD-interacting protein 2"
                     /protein_id="AAC00021.1"
                     /db_xref="GI:558844"
                     /translation="MEGVELKEEWQDEDFPIPLPEDDSIEADILAITGPEDQPGSLEV
                     NGNKVPKKLMAPDISLTLDPSDGSVLSDDLDESGEIDLDGLDTPSENSNEFEWEDDLP
                     KPKTTEVIRKGSIIEYIAAEEKEDGFPWFMFRIGEQDHRVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQPNYRYLMDNLFKYVIGTLELLVAENYMIVYLNGAITPE
                     KMPSLGWLPKCYQQIDPPLPKNLKSLIIVHPSWFIPTLLAVTPPFISSKFSQKIRYVF
                     NLAELAELVPMEYVGIPECIKQVDQELNGKQDEPKNEQ"
BASE COUNT      734 a    393 c    511 g    744 t
ORIGIN


  Query Match             6.9%;  Score 286.2;  DB 11;  Length 2382;
  Best Local Similarity   61.7%;  Pred. No. 2.4e-45;
  Matches  513;  Conservative    0;  Mismatches  303;  Indels   15;  Gaps    3;

Qy    199 gagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatctat 258
          ||||| || || | || ||| | || |   | ||||| | ||||| || ||| |
Db    358 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGCTCTGT 417

Qy    259 tctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttga 318
          | || |||||      || |||   |  | ||||| |  |||     ||
Db    418 ATTGTCAGATGATTTGGATGAAAGT---GGGGAGATTGACTTAGATGGCTTAGA---CAC 471

Qy    319 tacccccgatgaagcagattcttttgagtacactggccatgatcccacagccaacaaaga 378
          |  | ||  |     ||   |||||||     |   ||  ||| | | |||| |
Db    472 ACCGTCAGAGAATAGTAATGAGTTTGAGTGGGAAGATGATCTTCCAAAACCCAAGACTAC 531

Qy    379 ttctggccaagagtcagagtctattccagaatatacggccgaagaggaacgggaggacaa 438
```

```
Db    532 TGAACTAATTAGCAAAGGCTCAATTACTGATATATCAGTAATA                 591

Qy    439 ccggctttggatgacagtggtcattgaagaaaagaggcagccgattgacatcaaatccat 498
          |   |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    592 ACGACGCTGGCGTATGTTCAGGATTGGAGAATAGGACTACAACTTTGATATGAAGCAAT 651

Qy    499 cgagccctacaggagagtcatttctcagcgaaaagagttcgaatactatgaaaagttct 558
          || |||||  |  | |||  ||  |  |  |  |  |  |  |  |  |  |  |  |
Db    652 TGAACCCTATAAAAAAGTTATCAGCCATGG----------CTTTATTGTGTTCAATGATT 702

Qy    559 aaatgccatcattgtgtttgccgcctgttttctggccagacagcaactgaagcgaattacca 618
          |||||||||  |||| |||| |  |  |  |  |  |  |  |  |  |  |  |  |
Db    703 AAATGCCATTGTTGTATTTGCTGTCTGTTTTCATGCCTGAAAGTAGTCAGCGTAAATTATAC 762

Qy    619 ctatgtcatggaaaatcttttcctatatgtaatgagtacttttagagtttatgtttgtga 678
          ||  | ||||| |||||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    763 ATACCTGATGGACAATCTTTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTC 822

Qy    679 agactatatgattgtgtacttgaatggtgcagtccaagaagcgtggtgtccagggctaag 738
          | |||| |||||  || || ||  |  |  |  |  |  |  |  |  |  |  |  |
Db    822 AAACTACATGATAGTTTATTTAAATTGTTCAATAATTCTAATAAAAATCCTTAATTCTCC 882

Qy    739 ctggatgaagaagatgctaccagcttcattttcctgtgcgtgtttgatgtttgaattttgaaatcatt 798
          ||| | | |||||  || ||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    883 ATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGTTACTGAAAAAATCTAAAATCTCT 942

Qy    799 catcattgttcatccatcttgcttcatcagaagaatcctttgtctgtgacagcgaccttttat 858
          |||||||| ||||| ||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    943 AATCATTGTACATCCTTCTTGCTTTATCAGAATAGTTTCTGGCTCTTTAGAAGATTTTTAT 1002

Qy    859 aagttcaaaattcagcagtaaaattaaatatctcaatagcttatcagaactgaatcagcct 918
          || || ||||||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   1003 TAGCTCGAAATTCAGCCAAAAAATTAGATACGTGTTAATTTCGTAGAACTAGCAGAAGT 1062

Qy    919 gatcccaatggattgcatccacattcctagtgatcatcatcctcatttgcttgaatctgaa 978
          |||| |||| | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   1063 TGTCCCCATGGAATAGGTTGGCATAGCAGTATCATATTTTTTTTTTTTTTTTTTTTTTA 1122

Qy    979 agaaaagccttagttggccatgctggaagaagaggatgcttttctggttca 1029
          | |||| |   |    | ||     |||  |  |   ||| || ||
Db   1123 TGGAAAACAAGATGAACCGAAAAATGAACAGTAAGTTTGGCATCTAGTCCA 1175


RESULT   6
AF035207
LOCUS       AF035207     2213 bp    mRNA            PRI       28-NOV-1998
DEFINITION  Mus musculus Nip21 mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip21 gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1998) Functional Genomics, Incyte Corporation,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1. .2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit5 and D9Mit21"
     gene            1. .2213
                     /gene="Nip21"
```

```
     CDS             134. .1114
                     /gene="Nip21"
                     /note="associates with motifs of both the adenovirus E1B
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP21"
                     /protein_id="AAC04329.1"
                     /db_xref "GI:2911348"
                     /translation="MEGVPIKEEWQDEDFPIPLPEDDSIEADTLDGTDPDRQPGSLEV
                     NGNKVRKKLMAPDISLTLDPGEGSLWSDDLDEAGEVDLEGLDTPSENSDEFEWEDDLP
                     KPKTTEVIPKGSITEYTAIEFKGDGPRWRMFRIGEQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESGQPNYRYLMDNLFKYVIGTLELLVAENYMIIYLNGATIFP
                     KMPSLGWLRPCYQQIDPRLRKNLKSLIIVHPSWFIPTLLAVTPPFISSKFSQKIPYVF
                     NLAELAELVPMEYVGIPECIKQYEEEKFKKPQKPVDQELNGKQEPPKSEQ"
BASE COUNT      637 a    400 c    571 g    605 t
ORIGIN


  Query Match             6.7%;  Score 240;  DB 12;  Length 2213;
  Best Local Similarity   63.5%;  Pred. No. 9.9e-45;
  Matches  487;  Conservative    0;  Mismatches  265;  Indels   15;  Gaps    3;

Qy     200 agaatcaaagctcacagcccaaatatcaatcttctctggaccaaagtgaaggatctatt 259
           |:|  |:|| | ||||| | |||  |   | ||||| |   |||||| ||| |
Db     281 AGGAAGAAACTGATGGCCCCAGACATCAGCCTGACCCTGGATCCTGGTGAAGGTTCTCTG 340

Qy     260 ctctctgatgatgactttgaagatgtcgagattgaattgacatcaatgtggatgaacttgat 319
           || ||||||   |||||   | | |  ||||| | | |   ||
Db     341 TGGTCCGATGAT---TTGGATGAAGCTGGAGAGGTTGACCTGGAAGGCTTAGA---CACC 394

Qy     320 gtcctcgatgaagagattctttttgagtctagtggcatgatccacagcgcaacaaagat 379
           | | |  |||   ||||||||     |   ||  ||||| | |||| |    |
Db     395 CCATCAGAGAACAGTGATGAGTTTGAGTGGGAAGATGATCTTCCCAAACCCAAAACTACT 454

Qy     380 tctggccaaagagtcaaagtctattccagaatatacggccgcaggaggaacgggaggacaac 439
           | ||     |  ||  || ||| | |  |  || |||  ||| ||| ||  ||  ||
Db     455 GAAGTCATTAGGAAAAGGCTCGATTACTGAGTACACAGCCACAGAAGAAAAGGGCGATGGA 514

Qy     440 cggctttggatgacagtggttcattggaggaacaagagcagcggcattgacatgaaggtcatc 499
           || |  |||   |   |   |||||||| || || ||  |  | ||||||||||  |||
Db     515 CGCCGCTGGCGCAGTCTCAGGATTGGAGAGCAAGACCACAGGGTGGACATGAAGGCGATC 574

Qy     500 gaggcctacagggaagatcatttctcacggagtcgattcaggatactatgggacggctcta 559
           |||||||||| | ||| |||   ||  |        |||||| ||||||||||  ||
Db     575 GAGCCCTACAAAAAAGTTATTAGCCATGG---------AGGATATTATGGGGACGGCTTA 625

Qy     560 aatgccatcattgtgtttgccgctgttttctccagacagcagtcgggcggattaccac 619
           ||||||||  |||| |||| | ||||||    || || ||  ||| | |   | ||
Db     626 AAIGCCAITGTIGIATTTGCTGTCTGTTTCAITCCTGAGAGTGGTCAGCCTAACTATAGA 685

Qy     620 tatgtcatggaaaatctttcctatatgtaattagtactttagagttgatggtagctgaa 679
           |||| | | || || | |   |||||  |  | ||||||||| || |||||| ||
Db     686 IATTTGATGGACAAICICITTAAAIAIGTTAICGGCACTTTAGAGCTGTTGGTAGCAGAG 745

Qy     680 gacttatatgatttgtactttaaatgttgcaa tccaagaagtaggatgccaaggctagac 739
           |||| |||||  | ||||| |||||||| ||  |  |||| |  |||||  | | ||
Db     746 AACTACATGATCATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGA 805

Qy     740 tgaatgaaagaaatgctaccagatgattgacatacggttgagtaagaaatttgaaatcattc 799
           ||| | | | |||||||||  ||||  |  || || ||||| ||  | || ||  |
Db     806 IGGCICAGGAGATGCIACCAGCAAATTGATAGACGATTAAGCAAAAACCTTAAGTCTCTA 865

Qy     800 atcattgttcatccatcttggttcatcagaatgaatccttgctgtgacacggacctttttata 859
           |||||||| || || |||.|||| ||| .|||  | || ||||| |||||||| |||||
Db     866 AICATTGTGCACCCCTCTTGGTTTAICCGAATGACTTCTGGCTGTTACACGACCATTTATC 925

Qy     860 agttcaaaattcagcagtaaaattaaatatgtcaatagcttatcagaactcagtgggctg 919
           ||||| | ||||    || || | ||  ||  || |||  | || ||    | |||
Db     926 AGTTCAAAGTTCAGTCAGAAGATCAGATACTCTTTAACTTGGCCGAGCTTGGCAGAGCTT 985

Qy     920 atcccaatggattgcttcaacattccagaggaaatcatcaaatattga 966
           | || ||||| |   ||  ||| ||| ||  ||||   |   |||
Db     986 GTTCCCATGGAGTATGTCGGCATACCAGAGTTCIATCAAACAGTATGA 1032
```

```
RESULT   7
AC012249
LOCUS       AC012249    58386 bp    DNA             HTG       21-OCT-1999
DEFINITION  Homo sapiens clone 16_N_23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC012249
VERSION     AC012249.1  GI:6091795
KEYWORDS    HTG; HTGS_PHASE0
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone 16_N_23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58386)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,P., Boguslavsky,I., Bookhalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,T., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,S.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heeford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karlins,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McPherran,K., Melnikhin,I., Molnar,M.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,E.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vasilliev,N., Vu,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This record contains 78 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1      799: contig of 799 bp in length
            *                     gap of unknown length
            *       800     1536: contig of 737 bp in length
            *                     gap of unknown length
            *      1537     2265: contig of 729 bp in length
            *                     gap of unknown length
            *      2266     2972: contig of 707 bp in length
            *                     gap of unknown length
            *      2973     3725: contig of 753 bp in length
            *                     gap of unknown length
            *      3726     4475: contig of 750 bp in length
            *                     gap of unknown length
            *      4476     5247: contig of 772 bp in length
            *                     gap of unknown length
            *      5248     5992: contig of 745 bp in length
            *                     gap of unknown length
            *      5993     6730: contig of 738 bp in length
            *                     gap of unknown length
            *      6731     7474: contig of 744 bp in length
            *                     gap of unknown length
            *      7475     8210: contig of 736 bp in length
            *                     gap of unknown length
            *      8211     8928: contig of 718 bp in length
            *                     gap of unknown length
```

```
*     8929     9640: contig of 712 bp in length
*           gap of unknown length
*     9641    10387: contig of 747 bp in length
*           gap of unknown length
*    10388    11120: contig of 733 bp in length
*           gap of unknown length
*    11121    11879: contig of 759 bp in length
*           gap of unknown length
*    11880    12648: contig of 769 bp in length
*           gap of unknown length
*    12649    13441: contig of 793 bp in length
*           gap of unknown length
*    13442    14200: contig of 759 bp in length
*           gap of unknown length
*    14201    14936: contig of 736 bp in length
*           gap of unknown length
*    14937    15720: contig of 784 bp in length
*           gap of unknown length
*    15721    16404: contig of 684 bp in length
*           gap of unknown length
*    16405    17121: contig of 717 bp in length
*           gap of unknown length
*    17122    17847: contig of 726 bp in length
*           gap of unknown length
*    17848    18601: contig of 754 bp in length
*           gap of unknown length
*    18602    19316: contig of 715 bp in length
*           gap of unknown length
*    19317    20058: contig of 742 bp in length
*           gap of unknown length
*    20059    20787: contig of 729 bp in length
*           gap of unknown length
*    20788    21518: contig of 731 bp in length
*           gap of unknown length
*    21519    22254: contig of 736 bp in length
*           gap of unknown length
*    22255    23015: contig of 761 bp in length
*           gap of unknown length
*    23016    23737: contig of 722 bp in length
*           gap of unknown length
*    23738    24453: contig of 716 bp in length
*           gap of unknown length
*    24454    25182: contig of 729 bp in length
*           gap of unknown length
*    25183    25937: contig of 755 bp in length
*           gap of unknown length
*    25938    26683: contig of 746 bp in length
*           gap of unknown length
*    26684    27431: contig of 748 bp in length
*           gap of unknown length
*    27432    28179: contig of 748 bp in length
*           gap of unknown length
*    28180    28941: contig of 762 bp in length
*           gap of unknown length
*    28942    29667: contig of 726 bp in length
*           gap of unknown length
*    29668    30386: contig of 719 bp in length
*           gap of unknown length
*    30387    31105: contig of 719 bp in length
*           gap of unknown length
*    31106    31889: contig of 784 bp in length
*           gap of unknown length
*    31890    32643: contig of 754 bp in length
*           gap of unknown length
*    32644    33422: contig of 779 bp in length
*           gap of unknown length
*    33423    34169: contig of 747 bp in length
*           gap of unknown length
*    34170    34901: contig of 732 bp in length
*           gap of unknown length
*    34902    35643: contig of 742 bp in length
*           gap of unknown length
*    35644    36403: contig of 760 bp in length
*           gap of unknown length
*    36404    37119: contig of 716 bp in length
*           gap of unknown length
*    37120    37831: contig of 712 bp in length
*           gap of unknown length
*    37832    38555: contig of 724 bp in length
*           gap of unknown length
*    38556    39332: contig of 777 bp in length
*           gap of unknown length
*    39333    40130: contig of 798 bp in length
*           gap of unknown length
*    40131    40891: contig of 761 bp in length
*           gap of unknown length
*    40892    41629: contig of 738 bp in length
*           gap of unknown length
*    41630    42372: contig of 743 bp in length
*           gap of unknown length
*    42373    43156: contig of 784 bp in length
*           gap of unknown length
*    43157    43889: contig of 733 bp in length
*           gap of unknown length
*    43890    44610: contig of 721 bp in length
*           gap of unknown length
*    44611    45411: contig of 801 bp in length
*           gap of unknown length
*    45412    46177: contig of 766 bp in length
*           gap of unknown length
*    46178    46940: contig of 763 bp in length
*           gap of unknown length
*    46941    47711: contig of 771 bp in length
*           gap of unknown length
*    47712    48462: contig of 751 bp in length
*           gap of unknown length
*    48463    49258: contig of 796 bp in length
*           gap of unknown length
*    49259    50004: contig of 746 bp in length
*           gap of unknown length
*    50005    50781: contig of 777 bp in length
*           gap of unknown length
*    50782    51487: contig of 706 bp in length
*           gap of unknown length
*    51488    52199: contig of 712 bp in length
*           gap of unknown length
*    52200    52928: contig of 729 bp in length
*           gap of unknown length
*    52929    53736: contig of 808 bp in length
*           gap of unknown length
*    53737    54547: contig of 811 bp in length
*           gap of unknown length
*    54548    55305: contig of 758 bp in length
*           gap of unknown length
*    55306    56070: contig of 765 bp in length
*           gap of unknown length
*    56071    56822: contig of 752 bp in length
*           gap of unknown length
*    56823    57607: contig of 785 bp in length
*           gap of unknown length
*    57608    58386: contig of 779 bp in length

  Query Match              4.9%;  Score 215.2;  DB 42;  Length 58386;
  Best Local Similarity   98.2%;  Pred. No. 3.8e-33;
  Matches  219;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy    3969 gtagtttgttccctagaaacatttagatgtgacaattactttttgtttcttaaa 4022
           |||| ||||||||||||||| | | | | | | | | | | | | | | | |
Db   20090 GIAGGCIGTICCCTIAGAAACAIIIAIAIGIGACAATTAATTTTTTAIAIAIC-AAC 20148

Qy    4023 gttttctatcatgcattggattgctcagaataaagtgtcttttgaacttcgttttgctc 4082
           |||||||||||||||||||| | | | | | | | | | | | | | | | | |
Db   20149 GTTTTTCTAICAIGCAIIGGAI-G-CAGAAIAAAIG-G-GTTAGATTTTGTTTTATTA 20208

Qy    4083 aataaatttctccataatgtagatttaatgatatgaaagtcttttaatgaaaatatatc 4142
           |||| ||||||||| | ||| | | | | | | | | | | | | | | | |
```

```
Db  20269 AAIAAAIIIIIIAIAAIIIAGAIIAAIAAIA.IAAAGIGIIIAAIGACAGAATATAIIIA 20168

Qy   4143 tatagcctcactgtataattcagaaataaaaattgattctgca 4185
          |||||||||||||||||||||||||||| ||  |||||||| |
Db  20269 IAIAGCCICACIGIAIAAIICAGAAATAAAAA.IGATTCTGCA 20311


RESULT   8
G15207/c
LOCUS       G15207         217 bp    DNA             STS       04-JAN-1996
DEFINITION  human STS SHGC-15934.
ACCESSION   G15207
VERSION     G15207.1  GI:1131970
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata, Osteichthyes, Sarcopterygii, Choanata;
            Tetrapoda, Amniota, Mammalia, Theria, Eutheria, Archonta, Primates,
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157219687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: IGCAGAAICAAIIIIIAIIIIGA
            Primer B: AAGTGTCTGTTAGACTTCGIIITGG
            STS size: 130
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         36
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol.      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs derived from P22601--Merck/UniEST.
FEATURES             Location/Qualifiers
     source          1..217
                     /organism="Homo sapiens"
     STS             2..131
     primer_bind     2..25
     primer_bind     complement(107..131)
BASE COUNT       80 a     27 c     32 g     78 t
ORIGIN


  Query Match             4.8%;  Score 2[illegible].4;  DB 13,  Length 217,
  Best Local Similarity   99.1%;  Pred. No. 2e-29,
  Matches  215;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy   3971 ttccttgaaaacatttagatgtgcacaaattaatcttttatatatctaaaggtttttc 4029
          | ||||||||||||||||||||||||||||||'|| ||||||||||||||| |||||||||||||

Db    217 TCCTCTTAGAAACATTTAGATGTTCACAAATTAATCTTTTATATATCTAAAGGTTTTTC 158

Qy   4030 tatcatgcattggattgcttagaatgaagtgtctgttagacttcgttttgtgaaataaa 4089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 TATCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTAGACTTCGTTTTGTGAAATAAA 98

Qy   4090 tctccataatgtagattaataatataaaagtctttaatgacacaatatatctatataaaa 4149
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 TCTCCATAATGTAGATTAATAATATAAAAGTCTTTAATGACACAATATATCTATATAGTT 38

Qy   4150 tcactgtataattcagaaataaaaattgattctgcaa 4186
          ||||||||||'||||||||||||.|||||'|||||||
Db     37 TCACTGTATAATTCAGAAATAAAAATTGATTCTGCAA 1


RESULT   9
AC009556
LOCUS       AC009556   169798 bp    DNA             HTG       [illegible]-JAN-1999
DEFINITION  Homo sapiens chromosome 19, clone RP11-[illegible] [illegible] DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 19, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Berna,N., Beskeriy,R., Bowe,T., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,F., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,T.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Levy,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Myshchenkyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Parlin,E.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Savery,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6711453.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 126295 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000 agarose-fp
                Insert size: 169798 supercontigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
```

```
  Quality coverage: 4 1 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1016: contig of 1016 bp in length
*                   gap of unknown length
*     1017     2036: contig of 1020 bp in length
*                   gap of unknown length
*     2037     3371: contig of 1335 bp in length
*                   gap of unknown length
*     3372     4972: contig of 1601 bp in length
*                   gap of unknown length
*     4973     5981: contig of 1009 bp in length
*                   gap of unknown length
*     5982     8008: contig of 2027 bp in length
*                   gap of unknown length
*     8009     9286: contig of 1278 bp in length
*                   gap of unknown length
*     9287    11243: contig of 1957 bp in length
*                   gap of unknown length
*    11244    14312: contig of 3069 bp in length
*                   gap of unknown length
*    14313    18423: contig of 4111 bp in length
*                   gap of unknown length
*    18424    23065: contig of 4642 bp in length
*                   gap of unknown length
*    23066    30558: contig of 7493 bp in length
*                   gap of unknown length
*    30559    37537: contig of 6979 bp in length
*                   gap of unknown length
*    37538    44831: contig of 7294 bp in length
*                   gap of unknown length
*    44832    50944: contig of 6113 bp in length
*                   gap of unknown length
*    50945    58263: contig of 7319 bp in length
*                   gap of unknown length
*    58264    65216: contig of 6953 bp in length
*                   gap of unknown length
*    65217    72310: contig of 7094 bp in length
*                   gap of unknown length
*    72311    78799: contig of 6489 bp in length
*                   gap of unknown length
*    78800    87607: contig of 8808 bp in length
*                   gap of unknown length
*    87608    98337: contig of 10730 bp in length
*                   gap of unknown length
*    98338   112198: contig of 13861 bp in length
*                   gap of unknown length
*   112199   126330: contig of 14132 bp in length
*                   gap of unknown length
*   126331   140617: contig of 14287 bp in length
*                   gap of unknown length
*   140618   153843: contig of 13226 bp in length
*                   gap of unknown length
*   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


  Query Match             4.9%;  Score 202.4;  DB 44,  Length 169798,
```

```
  Best Local Similarity   99.5%;  Pred. No. 1.9e-29;
  Matches  203;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     158 gaacctcctaatttctctggatctttaatgaca.tcatcttggcaaatttaaa.tgaaaa 217
           |||||||||||||||||||| | | | | | | | | | | | | | | | | | | | |
Db  119825 GAACCTCCTAATTCTCTGGATCTTAATGA.ATTCATCCTTGGCAAATTAAATCAAGTTCACAGCC 119884

Qy     218 ccaaatatcaatctttctctggacaaagtgaaggatctatttctttctgatgatgaattg 277
           |||||||||||||||||||| | | | | | | | | | | | | | | | | | | | |
Db  119885 CCAAATATCAATCTTTCTCTGGACAAAGTGAAGGATCTATTTCTTTCTGATGATGAATTG 119944

Qy     278 gacagtccagatgaaattgacatcaatctggatgaacttgatatcccccgattaagcagat 337
           |||||||||||||||||||| | | | | | | | | | | | | | | | | | | | |
Db  119945 GACAGTCCAGATGAAATTGACATCAATCTGGATGAACTTGATATCCCCCGATTAAGCAGAT 120004

Qy     338 tcttttgagtacactggccatgat 361
           ||||||||||||||||||||||.
Db  120005 TCTTTTGAGTACACTGGCCATGGT 120028


RESULT  10
G32285/c
LOCUS       G32285          117 bp    DNA             STS       24-SEP-1999
DEFINITION  A005S17 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G32285
VERSION     G32285.1  GI:5922806
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   Adams,M.D.
  TITLE     Human STS sequences
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Email: mdadams@tigr.org
            Primer A: TGAATTATACAGTGAGGCTATAT
            Primer B: TCAGAATAAAGTGTCTGTTAGAC
            STS size: 117
            PCR Profile:
                      Denaturation:  96C 5min
                      Anneal:        54C 30sec
                      Extend:        72C 30sec
                      Denature:      96C 30sec
                      FinalExtend:   72C  5min
                      Cycles:        30
            Protocol:
                      GenomicDNA:    25 ng
                      Primer:         0.43 uM each
                      dNTPs:        200 uM each
                      AmpliTaq:       0.5 units
                      TaqStart Ab:    0.5 units
                      Total Volume: 10 ul

            Buffer:
                      Tris-HCl pH8.6: 100 mM
                      KCl:            500 mM
                      MgCl2:           20 mM
                      Triton X-100:    1%
                      Concentration:   1X

            Prepared with primer pairs derived from THC116446; GenBank
            Accession Numbers-  R45391, R45388, R65155, R16465, R228 1, N26684,
            R32435, D92189, R45390, AA850298, L59441.
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human"
```

```
     STS             1. .117
     primer_bind     1. .22
     primer_bind     complement(95. .117)
BASE COUNT       43 a     14 c     16 g     44 t
ORIGIN


  Query Match              2.8%;  Score 117;  DB 1;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-13;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4048 tcagaataaagtgtctgttagacttcgttttgctaaataaaattctccataatgtagatta 4107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117 TCAGAATAAAGTGTCTGTTAGACTTCGTTTTGCTAAATAAAATTCTCCATAATGTAGATTA 58

Qy   4108 ataatataaaagtctttaatgacacaatatatttatatagcctcactgtataattca 4164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57 ATAATATAAAAGTCTTTAATGACACAATATATTTATATAGCCTCACTGTATAATTCA 1


RESULT  11
AP000462/c
LOCUS       AP000462   114467 bp    DNA             PRI       20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 11q23, clone:B4H11, complete
            sequence.
ACCESSION   AP000462
VERSION     AP000462.1  GI:5881777
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:B4H11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114467)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 114,467 genomic DNA of 11q23
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 114467)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
FEATURES             Location/Qualifiers
     source          1. .114467
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="B4H11"
                     /map="11q23"
BASE COUNT    33083 a   20254 c   22317 g   38813 t
ORIGIN


  Query Match              2.8%;  Score 97.2;  DB 10;  Length 114467;
  Best Local Similarity   67.7%;  Pred. No. 2.6e-09;
  Matches  203;  Conservative    0;  Mismatches   78;  Indels   19;  Gaps    4;

Qy   2035 ctgcaagctaagaatggtttttacatttgggattggaaaaatttttttttgaatatttcat 2094
          ||| |||||| |||| || |||     |      | |||||| || ||  |||||| |||
Db  72188 CTGCAAGCTAAGAATAGTATTT---TTCAATTTTTGAAAAAATTTTTTAAGAATATATCAT 72179

Qy   2095 gacacatgaaa-------attattcaaaatgttagtgccgataaataaagtggtactgaaa 2147
          ||||  ||||       |  ||||||||    |||| | ||||||||||  || || ||
Db  72078 AACACCCGAAAAACACATAAAATTCAAATTGCAGTGTCCATAAATAAAGTTTTATTGGAA 72019

Qy   2148 cacagccacacaaaattgtttttgtactgtctacagctacttcacactacagccgcaga 2207
          || |||||  |  | | ||||   || ||||||||||  |||||||||||||| | |||||
Db  72018 CATAGCCATGCTCATTCGTTTACATATTGTCTACAGCAGCTTTCACACTACAACAGCAGA 71959

Qy   2208 gctgagcagtt---cagcagacctatttccacaatgcctaaaaattttttttttt   2263
          ||||||  ||||    |  |||||||||| |||||||  |||||| | ||| ||| |
Db  71958 GCTGAGTAGTTGTGAAAGAGACCGTATGCCCACAAAGCCTAACATATTTATTACTTTT 71899

Qy   2264 ----tacagaaaaagtttgctgt-----tgtctagcaaagcgtcttttttcgtcgtc 2318
               ||||||||||||||||  | |  |||| ||| | |  |  | | |  ||| |||
Db  71898 CCTCTACAGAAAAAGTTTGCCAA[C]TTTGTTATAGTAGAATAGAGACCTAAATTCCTAAAT 71839


RESULT  12
AC002981/c
LOCUS       AC002981   153568 bp    DNA             PRI       22-JAN-1998
DEFINITION  Homo sapiens Xp22 BAC GS279A12 (Genome Systems)  complete sequence.
ACCESSION   AC002981
VERSION     AC002981.1  GI:2804306
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153568)
  AUTHORS   Muzny,D., Aronson,A.D., Brundage,E., Carvalli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,T., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,I., Kampal,P.,
            Karpathy,S., Leal,B., Liu,W., Logan,G., Lu,J., Ly,T., Martinez,C.,
            Oswal,G., Perez,L., Rashid,N.D., Rowland,P., Savage,L.,
            Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q., Worley,K.C.,
            Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153568)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 153568)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 22, 1998 this sequence version replaced gi:2668524.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 1 reads with no
            ambiguities.  If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were shown if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .153568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GS279A12"
                     /chromosome="X"
     repeat_region   1205. .1287
                     /rpt_family="L1_CP_PLAM"
     repeat_region   1692. .1794
                     /rpt_family="CHARLIE1b"
     repeat_region   1800. .1913
                     /rpt_family="(TAA)n"
     repeat_region   1914. .2296
                     /rpt_family="CHARLIE1b"
     repeat_region   complement(2734. .2978)
                     /rpt_family="MIR"
```

```
repeat_region   complement(3929. .4230)
                /rpt_family="AluSx"
repeat_region   complement(4390. .4537)
                /rpt_family="MER20"
repeat_region   4681. .4984
                /rpt_family="L2"
repeat_region   5415. .5480
                /rpt_family="AT_rich"
repeat_region   5502. .5689
                /rpt_family="L2"
repeat_region   complement(8007. .8381)
                /rpt_family="THE1C"
repeat_region   complement(8382. .9957)
                /rpt_family="THE1C-internal"
repeat_region   complement(9959. .10322)
                /rpt_family="THE1C"
repeat_region   10534. .10645
                /rpt_family="MIR"
repeat_region   10877. .10956
                /rpt_family="(GA)n"
repeat_region   11278. .11386
                /rpt_family="L2"
repeat_region   12578. .12699
                /rpt_family="MIR"
repeat_region   complement(12951. .13061)
                /rpt_family="L2"
repeat_region   complement(14032. .14163)
                /rpt_family="FLAM_C"
repeat_region   complement(14219. .14336)
                /rpt_family="AluJo"
repeat_region   14417. .15229
                /rpt_family="L2"
repeat_region   complement(15737. .15801)
                /rpt_family="MIR"
repeat_region   complement(16977. .17463)
                /rpt_family="L2"
STS             17651. .17827
                /standard_name="AFM270xf5, Chr. X, Homo sapiens, DXS7104"
                /db_xref="dbSTS:47221"
repeat_region   17684. .17730
                /rpt_family="(CA)n"
repeat_region   complement(18599. .18631)
                /rpt_family="(CAAAA)n"
repeat_region   19445. .19487
                /rpt_family="GC_rich"
repeat_region   complement(24029. .24340)
                /rpt_family="MER58B"
repeat_region   25553. .25590
                /rpt_family="(TA)n"
repeat_region   complement(26396. .26534)
                /rpt_family="MER58A"
repeat_region   26875. .27177
                /rpt_family="AluY"
repeat_region   27239. .27291
                /rpt_family="(CA)n"
repeat_region   27341. .27492
                /rpt_family="L2"
repeat_region   28459. .29916
                /rpt_family="TIGGER1"
repeat_region   29919. .30029
                /rpt_family="HY1"
repeat_region   30035. .30975
                /rpt_family="TIGGER1"
repeat_region   31504. .31623
                /rpt_family="TIGGER1"
repeat_region   complement(32185. .32205)
                /rpt_family="AT_rich"
repeat_region   complement(32565. .32692)
                /rpt_family="L1ME"
repeat_region   complement(33718. .34021)
                /rpt_family="AluYb8"
repeat_region   complement(34552. .34859)
                /rpt_family="MER58B"
repeat_region   36375. .36436
                /rpt_family="POLY_A"
repeat_region   complement(36582. .36724)
                /rpt_family="MER57A"
repeat_region   36884. .37044
                /rpt_family="MER2"
misc_feature    complement(37564. .37749)
                /note="Region: Similar to: M28 7' EST 1-3 Hippocampus,
                Stratagene (cat. #937135) Homo sapiens cDNA clone HHCH"
STS             37723. .37746
                /standard_name="DXS1[illegible]6E"
repeat_region   complement(37834. .38782)
                /rpt_family="(CA)n"
repeat_region   complement(38856. .38946)
                /rpt_family="L1ME1"
repeat_region   complement(41662. .41717)
                /rpt_family="(CA)n"
repeat_region   complement(41845. .41922)
                /rpt_family="(CAA)n"
repeat_region   complement(42057. .42251)
                /rpt_family="AluSx"
repeat_region   complement(42268. .42468)
                /rpt_family="MIR"
repeat_region   complement(44141. .44179)
                /rpt_family="(CAAA)n"
repeat_region   44256. .44294
                /rpt_family="POLY_A"
repeat_region   complement(44795. .44945)
                /rpt_family="POLY_A"
repeat_region   complement(44855. .48228)
                /rpt_family="L1MA2"
STS             46221. .46346
                /standard_name="STS[illegible] SHGC-21864-RA/SHGC-21[illegible] 7  Homo
                sapiens"
                /db_xref="dbSTS:7421"
repeat_region   48744. .48836
                /rpt_family="MER2"
repeat_region   complement(49417. .49716)
                /rpt_family="AluY"
repeat_region   complement(50267. .50483)
                /rpt_family="MER57A"
repeat_region   50670. .50744
                /rpt_family="MER57A"
repeat_region   complement(53275. .53438)
                /rpt_family="MER57A"
repeat_region   54868. .55164
                /rpt_family="MER57B"
repeat_region   55427. .55542
                /rpt_family="MER58B"
repeat_region   55793. .55875
                /rpt_family="L2"
repeat_region   56203. .56632
                /rpt_family="MER57A"
repeat_region   complement(56311. .56436)
                /rpt_family="AluSq"
repeat_region   56508. .56670
                /rpt_family="MER57A"
repeat_region   complement(56884. .56955)
                /rpt_family="MER58B"
repeat_region   56757. .56974
                /rpt_family="MER58B"
repeat_region   complement(57[illegible]. .57[illegible])
                /rpt_family="MER57B"
repeat_region   complement(57[illegible]. .57[illegible])
                /rpt_family="AT_rich"
repeat_region   complement(57[illegible]. .57[illegible])

Query Match             2.2%;  Score 92.8;  DB 10;  Length 15[illegible];
Best Local Similarity   60.2%;  Pred. No. 1.8e-[illegible];
Matches  121;  Conservative    [illegible];  Mismatches  [illegible];  Indels  [illegible];  Gaps  4;

Qy   2022 gtctgcctcaggactgcaggctcaggctgtttcctcttttttcttttttttttttttttttttttttt 2081
```

```
Db  55137 GTTTTGATACCATTCACAAGCTAAGGATGGT...TTACACTTTTAATTGGTTGAAAAAATA 55078

Qy   2092 ttgaatatttcatgacacatgaaaattattcaaatgttagtaccgataaataaagtggta 2141
          |||||||||||||||||||| | | || || | ||| | |||| ||||| ||
Db  55077 AT-AATATTTCATGACACATGAAAGTAATTCAAATTTCAGTGTCTGTAAAAAAAGTTTTA 55019

Qy   2142 ctgaaaacacaaccaacacaaacttgttttttgtactgtctacagctactttcacactacagc 2201
          ||| |||||||| | || ||| || |||  |   ||| ||||||||||||||
Db  55018 CTGGAACACAGC--CCCATACTCATTCATTTATAAATTGTGGCTGCTTTCACACTACATG 54961

Qy   2202 cgcagagctgagcagttt--cagcagacagtatgtcccacaatgcctaaaacattgacta 2259
          ||||| ||| |||| | |||| | || |||| || ||  |||| | ||||
Db  54960 GGCAGAGTTGAATAGTTGTGAAAGAGACCATA|GGCCCAAAAAGCTGAAAATGTGTACTA 54901

Qy   2259 tg---  -tttacagaaaaaagtttgctgacccctgctcta 2292
          ||   |||||| ||| |||| ||     ||||
Db  54900 TGTGGCCTTTGGCAGAAGATGTTGGCTGACCTCCTTTCTA 54861


RESULT  13
AC011488/c
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *                   gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *                   gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *                   gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *                   gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *                   gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *                   gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *                   gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *                   gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *                   gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *                   gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *                   gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *                   gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *                   gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *                   gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *                   gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *                   gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *                   gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *                   gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *                   gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *                   gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *                   gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *                   gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *                   gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *                   gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *                   gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *                   gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *                   gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *                   gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *                   gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *                   gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *                   gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *                   gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *                   gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *                   gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *                   gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *                   gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *                   gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *                   gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *                   gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *                   gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *                   gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *                   gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *                   gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *                   gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *                   gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *                   gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *                   gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *                   gap of unknown length
            *    41840    42836: contig of 997 bp in length
```

```
            *                    gap of unknown length
            *    42837    43913. contig of 1077 bp in length
            *                    gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *                    gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *                    gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *                    gap of unknown length
            *   112601   179301. contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  40598 t     85 others
ORIGIN


  Query Match              2.0%;  Score 93;  DB 41;  Length 179301;
  Best Local Similarity   69.7%;  Pred. No. 2.5e-09;
  Matches  134;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    156 ctgaaaatatcaatctttttctggaccaaagtaaaggatctattctctctgatgataact 215
          |  | | |   | ||  || | ||     .|||| |  ||| |  | |||    |
Db 124306 CAGTTCCTTCCCAACACGTTAAATTTCAACCCTGCGCATCGTAAGAGGAAGACGCTGGTGG 124247

Qy    216 cccaaatatcaatctttttctggaccaaagtaaaggatctattctctctgatgataact 275
          ||||| | |||||  ||||||||||| || ||||| || ||  | || || |||||   ||
Db 124246 CCCCAGAGATCAACATTTTTCTCGGATCAGAGTGAGGGTCCCTCGTCGCATGATTTTT 124187

Qy    276 tgaaacgtccagaatgaaattgacatcaatgttgatgaacttgatacccccgatgaagcag 335
          ||| |  | ||||| | || || || ||||||||  | || ||||||||||||  | |
Db 124186 TGGATACTTTTGATGACTTGGATATTAACTTTGATGACATCGAGACCCCCGATGAGACCG 124127

Qy    336 attcttttgaatcactggccatg 359
          || |  | | ''''  ||  |''|
Db 124126 ACTTGCTGCAGTTCCTGGGGAATG 124103


RESULT 14
AC006963
LOCUS       AC006963    88215 bp    DNA             PRI       22-OCT-1999
DEFINITION  Homo sapiens PAC clone DJ0278L01 from X, complete sequence.
ACCESSION   AC006963
VERSION     AC006963.4  GI:4753290
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 88215)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 88215)
  AUTHORS   Holmes,A., Wohldmann,P., Tucci,S. and Merry,B.
  TITLE     The sequence of Homo sapiens PAC clone DJ0278D01
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 88215)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 88215)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 88215)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri, 63108, USA
COMMENT     On May 5, 1999 this sequence version replaced gi:4758296.
            SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once.  It could be larger if we picked a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequence 3 with an alternate
            chemistry, or sequenced by [illegible] (Phrap quality >=
            30); an attempt was made to resolve all sequence ambiguities, such
            as compressions and repeats. All regions were covered by sequence
            from more than one clone and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the chromosome X mapping group
            at the Sanger Centre, [illegible], Hinxton, UK.
            Further information can be found at
            http://www.sanger.ac.uk/X/

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced is [illegible]. A [illegible] start of
            this clone is at base position 1 of [illegible].
FEATURES             Location/Qualifiers
     source          1..88215
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="DJ0278L01"
                     /clone_lib="RPCI-1"
     STS             21..538
                     /db_xref="GI:[illegible]"
     repeat_region   462..698
                     /rpt_family="MIR"
     repeat_region   719..920
                     /rpt_family="MER1_type"
     repeat_region   1316..1670
                     /rpt_family="L2"
     repeat_region   1904..2061
                     /rpt_family="MIR"
     repeat_region   2091..2111
                     /rpt_family="AT_rich"
     repeat_region   2142..2273
                     /rpt_family="MIR"
     repeat_region   2274..2646
                     /rpt_family="MER1B"
     repeat_region   2647..2731
                     /rpt_family="MIR"
```

```
repeat_region   2974. .3096
                /rpt_family="MIR"
repeat_region   4179. .4413
                /rpt_family="MIR"
repeat_region   4634. .4715
                /rpt_family="L1"
repeat_region   5148. .5448
                /rpt_family="Alu"
repeat_region   5477. .5700
                /rpt_family="MER1_type"
repeat_region   5823. .5999
                /rpt_family="MER2_type"
repeat_region   5946. .6011
                /rpt_family="MER2_type"
repeat_region   6010. .6562
                /rpt_family="MaLR"
repeat_region   8333. .8688
                /rpt_family="L1"
repeat_region   9295. .9412
                /rpt_family="MIR"
repeat_region   9796. .9922
                /rpt_family="MIR"
repeat_region   10206. .10294
                /rpt_family="MIR"
repeat_region   11444. .12394
                /rpt_family="L1"
repeat_region   12395. .12746
                /rpt_family="MaLR"
repeat_region   12747. .14693
                /rpt_family="L1"
repeat_region   14694. .15055
                /rpt_family="MaLR"
repeat_region   15056. .15399
                /rpt_family="L1"
repeat_region   15409. .21426
                /rpt_family="L1"
repeat_region   21441. .22630
                /rpt_family="L1"
repeat_region   22631. .22916
                /rpt_family="Alu"
repeat_region   22917. .22943
                /rpt_family="(CAA)n"
repeat_region   22955. .26162
                /rpt_family="L1"
repeat_region   27338. .27700
                /rpt_family="L1"
repeat_region   29461. .29653
                /rpt_family="MER1_type"
repeat_region   30139. .30472
                /rpt_family="MER2_type"
repeat_region   30959. .31081
                /rpt_family="MaLR"
repeat_region   31563. .31768
                /rpt_family="L1"
repeat_region   34255. .34339
                /rpt_family="(TA)n"
repeat_region   34469. .34629
                /rpt_family="MER1_type"
repeat_region   34808. .34833
                /rpt_family="MER1_type"
repeat_region   35741. .35872
                /rpt_family="L2"
repeat_region   35985. .36215
                /rpt_family="MER1_type"
repeat_region   37803. .37892
                /rpt_family="MaLR"
repeat_region   38113. .38456
                /rpt_family="MaLR"
misc_feature    39216. .39738
                /note="similar to Mus musculus EST AI131585 (NID:g3601601)
                ue98f12.x1"
misc_feature    39216. .39448
                /note="match to EST N22179 (NID:g1128313) yw30d10.s1"
misc_feature    39216. .39739
                /note="match to EST AA131585 (NID:g1692467) ue98f12.s1"
misc_feature    39216. .39672
                /note="match to EST AI339007 (NID:g4075494) qg34h04.x1"
misc_feature    39217. .39663
                /note="match to EST N70642 (NID:g1227222) za85a11.s1"
misc_feature    39240. .39717
                /note="match to EST AA434479 (NID:g2139492) aa54b05.s1"
misc_feature    39240. .39480
                /note="similar to Rattus sp. EST AI008854 (NID:g4121886)"
misc_feature    39270. .39509
                /note="similar to Oryctolagus cuniculus EST D82452
                (NID:g3061409)"
repeat_region   39583. .39604
                /rpt_family="AT_rich"
misc_feature    39748. .40265
                /note="similar to Mus musculus EST AA183744 (NID:g1767174)
                mt20h06.r1"
misc_feature    39860. .40178
                /note="match to EST AA100396 (NID:g1646733) zn49e03.r1"
gene            complement(40211. .86977)
                /gene="WUGSC:H_DJ0278D01.1"
CDS             complement(join(40211. .40738,42836. .43344,50764. .51499,
                60924. .61120,81641. .81773,86827. .86977))

  Query Match             2.1%;  Score 99.6;  DB 40;  Length 78170;
  Best Local Similarity   63.8%;  Pred. No. 7.5e-08;
  Matches  153;  Conservative    0;  Mismatches   84;  Indels    3;  Gaps    1;

Qy    2071 aaaaatttttttttgaatatttcatgacacatgaaaattattcaaatgttaattaaaatga 2130
           ||| |    | |||  || | ||||| |  | |    | ||||||  |  |  | |
Db   35996 AAAGAAGAATATTGTGTAATACATGAAAATTATATGAAATTCAAATTTCAATTTTATAA 36055

Qy    2131 ataaagtggtactgaaacacagccacacaaacttgttttttatactatttaaaatactt 2190
           |||||||  || |||||||||| |||||  | |  |||  |  |||
Db   36056 ATAAAGTTTTATTGAAACACATCCACACTCATTCATTTACATATTGTCTGTTTTCTGCCTT 36115

Qy    2191 cacactacagccgcagagctgag                                    2247
           |||||| || | |||||  |||| ||||     | ||||| | || | ||||
Db   36116 CACACTGCAACAGCAGACATGAGTAGTTGTGATGAAGACCATATGCCCCATAAAAATTAA 36175

Qy    2248 aacattgactatgtttacagaaaaaagtttgctgacccctgctctagcaaaaaaaatcctt 2307
           || |||  ||   | |  | | |||||||||  | ||||| | ||  |  | |    ||
Db   36176 AATATTTGCTGGCTCTTTACAGAAACTTTGCCAAGCCCTGTTGTAATATCCTTTTTTTT 36235


RESULT  15
AC015396/c
LOCUS       AC015396    78170 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces
ACCESSION   AC015396
VERSION     AC015396.1  GI:6435939
KEYWORDS    HTG; HTGS_PHASE2
SOURCE      fruit fly
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78170)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CFM 10029602 by the submitter.
            For further information on this sequence e-mail to ...@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .78170
                     /organism="Drosophila melanogaster"
```

/db_xref="taxon:7227"
BASE COUNT 21762 a 17118 c 17312 g 21978 t
ORIGIN

Query Match 2.0%; Score 85.6; DB 43; Length 78170;
Best Local Similarity 60.5%; Pred. No. 4.4e-07;
Matches 161; Conservative 0; Mismatches 99; Indels 6; Gaps 1;

```
Qy     482 attgacatgaaggtcatcgagccctacaggaagagtcatttctcacggaggagattcagga 541
           || |||||| |||||| |||||||||| ' | || | || ||||| ||  ||
Db    8691 ATCGACATGCGGGTCATAGAGCCCTACAAGCCTGTGCTGTCCCACGGCGGCTAT------ 9637

Qy     542 tactatggggacggtctaaatgccatcattgtgtttgccgcctgttttctgccagacagc 601
             | |    | |||||  || |||||  |  | |||  |||||||   |||||  || ||
Db    8637 CTCAAATCTGGCGGTCAGAACGCCATTGTCATCTTTTGCGCCTGTCACCTGCCGGATAGG 8578

Qy     602 agtcgggcggattaccactatgtcatggaaaatcttttcctatatgtaataagtacttta 661
             |||| || |||  ||| || ||||| || | || || || ||  | |  ||  |
Db    8577 TCGAGGGCCGACTACAGTACGTGATGGACAATTTGTTTCTGTACGTGGTCAAGACCCTG 8518

Qy     662 gagttgatggtagctgaagactatatgattgtgtacttgaatggtgcaaccccaagaagg 721
           ||   | |'||  '|||| |||||| || |  ' ||  || |||| |   ||   ||  |
Db    8517 GAACAGTTGGTCACTGATGACTATGTGCTAATTCTGCATGGCGGCTCCAATAGGCGT 8458

Qy     722 aggatgccagggctaggctggatgaa 747
           '   |||||    |   |||| | ||
Db    8457 AATGTGCCACCTTTTCCCTGGCTTAA 8432
```

Search completed: February 9, 2000, 07:52:19
Job time: 12453 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 20:41:06 ; Search time 284.47 Seconds
(without alignments)
849.599 Million cell updates/sec

Title: US-09-174-937-6
Perfect score: 966
Sequence: 1 atggaggaggagacagagttt.....tgaagctgaaagaaaagcct 966

Scoring table: IDENTITY_NUC

Searched: 311895 seqs, 125096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	326.6	33.8	1227	1	X39873	Gastric cancer ass
2	326.6	33.8	916	1	X39970	Gastric cancer ass
3	294.8	30.5	1096	1	T09229	Apoptosis-regulati
4	294.8	30.5	1396	1	X84120	Human Nip2 encodin
C 5	35.2	3.6	110000	1	V21209_09	Continuation (7 of
6	34.2	3.5	1700	1	X29771	L lactis HsdM gene
7	33.2	3.4	9479	1	Q81792	B. subtilis biotin
C 8	33	3.4	777	1	V23857	Plant LAC enzyme D
C 9	33	3.4	110000	1	V21209_03	Continuation (4 of
C 10	32.8	3.4	10592	1	T75610	Canine herpesvirus
C 11	32.8	3.4	5495	1	T75590	Canine herpesvirus
12	32.8	3.4	910	1	V04600	Flea serine protea
C 13	32.8	3.4	5495	1	V66900	Canine herpes viru
14	32.8	3.4	5495	1	V66901	Canine herpes viru
C 15	32.8	3.4	1248	1	V66905	Canine herpes viru
C 16	32.8	3.4	10592	1	V66938	Canine herpes viru
17	32.8	3.4	10592	1	V66939	Canine herpes viru
C 18	32.6	3.4	3331	1	N91248	Sequence of pLH1 c
C 19	32.6	3.4	110000	1	X20249_09	Continuation (7 of
C 20	32.6	3.4	111309	1	X00250	Borrelia burgdorfe
21	32.4	3.4	1076	1	V08817	Gene No. 7 encodin
C 22	32.4	3.4	9432	1	X20291	Borrelia burgdorfe
C 23	32.4	3.4	110000	1	X20248_09	Continuation (5 of
C 24	32.4	3.4	110000	1	X20248_10	Continuation (6 of
25	32.4	3.4	110277	1	X20249	Borrelia burgdorfe
C 26	32.2	3.3	2379	1	M13400	Enterococcus faeca
C 27	32	3.3	472	1	T03441	Human matrix metal
28	32	3.3	1918	1	I45650	Mouse ST-S1 cDNA.
C 29	31.8	3.3	3456	1	I03436	Human matrix metal
C 30	31.8	3.3	3437	1	T03437	Human matrix metal
31	31.8	3.3	2277	1	V05270	Human telomerase p
32	31.8	3.3	2277	1	V13834	Homo sapiens ambig
33	31.6	3.3	1357	1	Q04493	Sequence encoding
34	31.6	3.3	2111	1	V24142	Homo sapiens BARD1
35	31.6	3.3	110000	1	X20248_10	Continuation (7 of
36	31.6	3.3	111309	1	X20250	Borrelia burgdorfe
37	31.4	3.3	779	1	V75627	Staphylococcus aur
38	31.4	3.3	2658	1	V80061	Aspartate-specific
C 39	31.2	3.2	4065	1	Q11540	Human Stem Cell Fa
C 40	31.2	3.2	521	1	T14487	Human EST genomic
C 41	31.2	3.2	11478	1	T74744	P... potato
C 42	31.2	3.2	11478	1	T74745	P... potato
C 43	31.2	3.2	1115	1	V04441	DNA ... a rib
C 44	31.2	3.2	11475	1	V64474	P... starch bran
C 45	31.2	3.2	11478	1	V74...	P... ...B sta

ALIGNMENTS

RESULT 1
X39873
ID X39873 standard; DNA; 1227 BP.
AC X39873;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated gene.
KW Cancer associated antigen; diagnosis; human; gastric cancer;
KW breast cancer; colon cancer; gastric cancer; ss.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, Tata Y, Old L;
PI Pfreundschuh M, Sahin U, Stenner M, Stockert E;
PI Tureci O;
DR WPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients useful in diagnosis
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67; Page 579-580; 787pp; English.
CC The invention relates to a method for diagnosing a cancer characterised
CC by expression of a human cancer associated antigen protein recognised by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM or expression product of the NAM; and (b) determining the
CC product complexed with an RNA molecule; and (c) determining the
CC interaction between the agent and the NAM in the sample to diagnose the
CC determination of the disorder. The agent is an antibody used in
CC the diagnosis, monitoring, research or treatment of cancers
CC characterised by the expression of gastric cancer associated antigens
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen proteins expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer and
CC lung cancer.
SQ Sequence 1227 BP; 388 A; 271 C; 334 G; 274 T;

Query Match 33.8%; Score 326.6; DB 1; Length 1227;
Best Local Similarity 94.4%; Pred. No. 1.3e-24;
Matches 872; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

Qy 1 atggaggaggagacagagtttcctgaaatcgcgtagtaccatatcatcaaaaaatgaatgtta 60
Db 825 ATGGAGGAGGAGACAGAGTTCCTGAAATCGCGAACTACTCATCATCAAAATGAATGTTA 884

Qy 61 ctgtcagaggatgtaggaatgggagctgtcgttgtgcccagaaaatcctgttttgtgtgtat 120
Db 885 CTGTCAGAGGATGTAGGAATGGAGCTGTTCGTTGAAGAGCTTTTCGTCAGTTTCATTTGT 944

Qy 121 gcagacatgaggccttaaatcttgattctgttgatgtgaaaatgatgatgacagatgtgtca 180
Db 945 GCAGACATGAGGCCTAAATCTTGATTCTGTTGATGTGAAAATGATGATGACAGATGTGTCA 1004

Qy 181 atcaagctcacagaggcgaaatcagtgcttgaagtgcaaaaaaaaaaaaaaaaaaaaaaaa 240

```
Db   1005 ATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTC 1064

Qy    241 tctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaacttgatac 299
          |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db   1065 TCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATAC 1124

Qy    300 ccccgatgaagcagattcttttgagtacactgccc---atgatcccacagccaacaaag 355
          |||||||||||||||||||||||||||||||||   ||| ||    | ||  ||     | ||||||
Db   1125 CCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCCCATGAAGAATCCCACAGCCACAAAG 1184

Qy    356 attctggcc--aagagtcagagtctattccagaa 387
          |||||||||  |  |||||||||||||| | | |
Db   1185 ATTCTGGCCCAAGAAGTCAGAGTCTATTTCNGGA 1218


RESULT   2
X39876
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer, colon cancer, gastric cancer, renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; 014679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The kits and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens
CC   the invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    209 G;    184 T;


  Query Match            33.8%;  Score 327.6;  DB 1;  Length 816;
  Best Local Similarity   94.4%;  Pred. No. 1e-92;
  Matches  372;  Conservative    0;  Mismatches   15;  Indels    7;  Gaps    3;

Qy      1 atggaaggaggagacagagttccttgagctcgtaaccaggatatcaagaccaaaatggacta 60
          |||||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db    414 ATGGAGGAGGAGACAGAGTTCCTTGAGCTCGTAACCAGGATATCAAGACCAAATGGACTA 473

Qy     61 ctatcagaggatgtaggaatggacatccccttttgaagagggcgtgctgagtcccagtgct 120
          |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||
Db    474 CTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCAGTGCT 533

Qy    121 gcagacatgaggcctgaacctcctaattctctggatcttaatgacactgatcctcttata 180
          |||||||||||||||||||||||| | | | |||||||||  | ||
Db    534 GCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTGATCCTCTTATA 593

Qy    181 atcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatctattctc 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    594 ATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTC 653

Qy    241 tctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaacttgatac 299
          |||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Db    654 TCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATAC 713

Qy    300 ccccgatgaagcagattcttttgagtacactgccc---atgatcccacagccaacaaag 355
          |||||||||||||||||||||||||||||||||   ||| ||    | ||  ||     | ||||||
Db    714 CCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCCCATGAAGAATCCCACAGCCACAAAG 773

Qy    356 attctggcc--aagagtcagagtctattccagaa 387
          |||||||||  |  |||||||||||||| | | |
Db    774 ATTCTGGCCCAAGAAGTCAGAGTCTATTTCNGGA 807

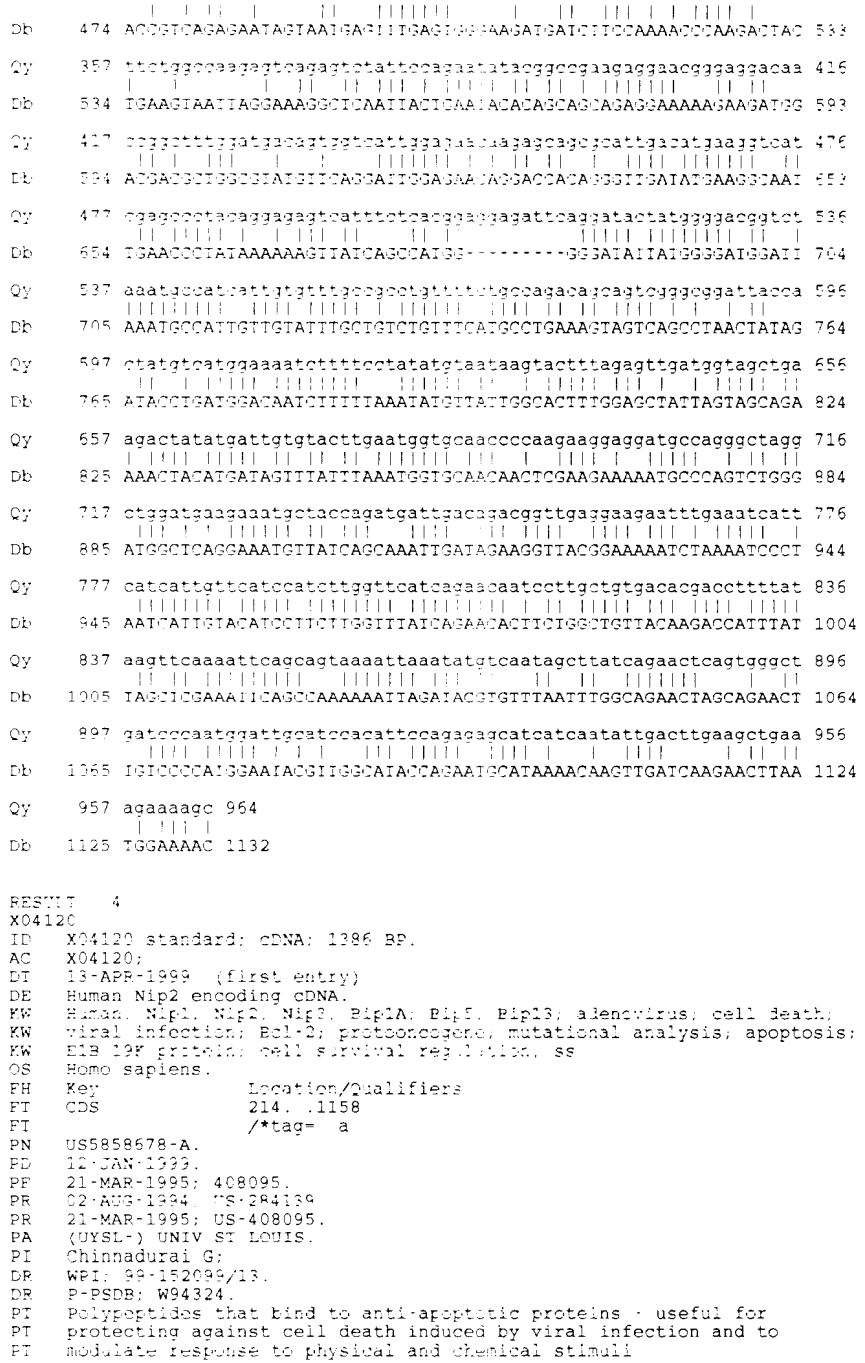

RESULT   3
T38228
ID   T38228 standard; cDNA; 1386 BP.
AC   T38228;
DT   18-DEC-1996  (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA
KW   Apoptosis-regulating protein; Nip2; adenovirus E1B 19K protein;
KW   Bcl-2; oncogene; cell death, cancer, tumour, immune disorder;
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214. .1158
FT                   /*tag= a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998.
PT   Nucleic acids encoding apoptosis regulating proteins - useful for
PT   diagnosing and treating immune disorders, malignancies, etc.
PS   Claim 1; Page 39-41; 60pp; English.
CC   6 cDNA clones (T38227-32) respectively code for novel apoptosis
CC   regulating proteins Nip1, Nip2, Nip3, BipIA, Bip5 and Bip6
CC   (W00997-W01002). Nip clones were isolated from human mRNA
CC   libraries using a yeast two-hybrid strategy with adenovirus E1B
CC   19K protein (W01010) as bait; Bip clones were isolated using
CC   Bcl-2 oncoprotein (W01018) as bait. E1B 19K and Bcl-2 both protect
CC   cells against apoptosis. Nip1, Nip2, Nip3 bind to 19K, BipIA and
CC   Bip5 bind to 19K and Bcl-2, and Bip5 binds to Bcl-2 alone. Nucleic
CC   acids can be used in diagnostic assays, to produce recombinant
CC   apoptosis-regulating proteins or to regulate cell death by
CC   increasing or suppressing expression of an apoptosis-blocking gene
CC   such as bcl-2.
SQ   Sequence   1386 BP;    428 A;    249 C;    336 G;    373 T;


  Query Match            29.5%;  Score 284.8;  DB 1;  Length 1386;
  Best Local Similarity   62.9%;  Pred. No. 1.8e-79;
  Matches  496;  Conservative    0;  Mismatches  277;  Indels   15;  Gaps    3;

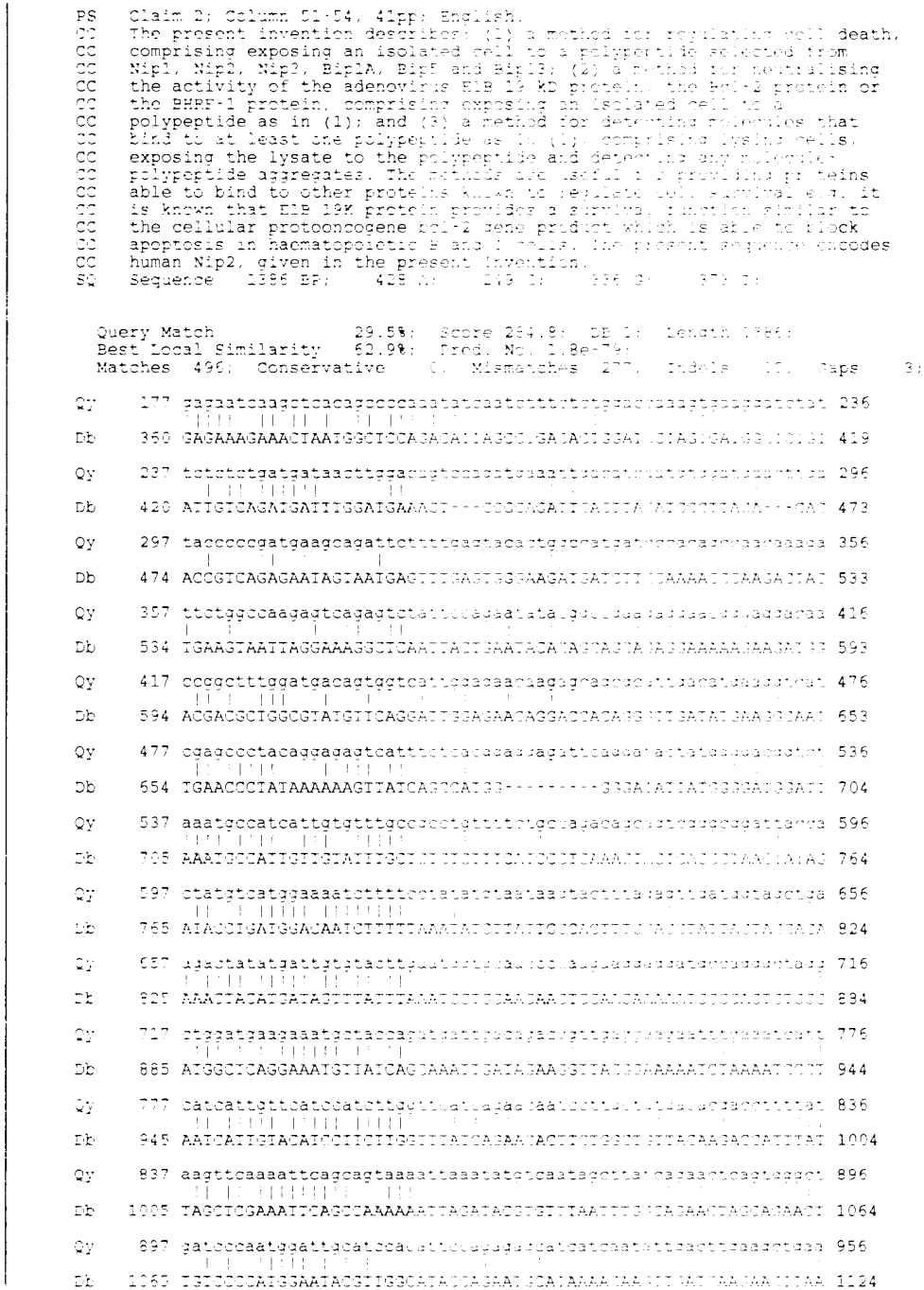
Qy    177 gagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatctat 236
          |||||  || || |  || |||    | |  || | ||||  |  |  |  |  |  |
Db    360 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGTTTTGT 419

Qy    237 tctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgat 296
          |  || ||||||     || |||   |  || |||||| |  |||    ||
Db    420 ATTGTCAGATGATTTGGATGAAAGT---GGGGAGATTGATTTAGATTTTTTAGA---TAT 473

Qy    297 taccccccgatgaagcagattctttttgagtacactgggccatgatgaagaatcccacagc 356
```



RESULT 4
X04120
ID X04120 standard; cDNA; 1386 BP.
AC X04120;
DT 13-APR-1999 (first entry)
DE Human Nip2 encoding cDNA.
KW Human; Nip1; Nip2; Nip3; Bip1A; Bip5; Bip13; adenovirus; cell death;
KW viral infection; Bcl-2; protooncogene; mutational analysis; apoptosis;
KW E1B 19K protein; cell survival regulation; ss
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 214. .1158
FT /*tag= a
PN US5858678-A.
PD 12-JAN-1999.
PF 21-MAR-1995; 408095.
PR 02-AUG-1994; US-284139
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 99-152099/13.
DR P-PSDB; W94324.
PT Polypeptides that bind to anti-apoptotic proteins - useful for
PT protecting against cell death induced by viral infection and to
PT modulate response to physical and chemical stimuli



PS Claim 2; Column 51-54, 41pp; English.
CC The present invention describes: (1) a method for regulating cell death,
CC comprising exposing an isolated cell to a polypeptide selected from
CC Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13; (2) a method for neutralising
CC the activity of the adenovirus E1B 19 kD protein, the bcl-2 protein or
CC the BHRF-1 protein, comprising exposing an isolated cell to a
CC polypeptide as in (1); and (3) a method for detecting molecules that
CC bind to at least one polypeptide as in (1), comprising lysing cells,
CC exposing the lysate to the polypeptide and detecting any molecular
CC polypeptide aggregates. The methods are useful in providing proteins
CC able to bind to other proteins known to regulate cell survival e.g. it
CC is known that E1B 19K protein provides a survival function similar to
CC the cellular protooncogene bcl-2 gene product which is able to block
CC apoptosis in haematopoietic B and T cells. The present sequence encodes
CC human Nip2, given in the present invention.
SQ Sequence 1386 BP; 428 A; 249 C; 306 G; 373 T;

Query Match 29.5%; Score 284.8; DB 1; Length 1386;
Best Local Similarity 62.9%; Prod. No. 1.8e-79;
Matches 496; Conservative 0; Mismatches 277; Indels 10; Gaps 3;

```
Qy     957 agaaaagc 964
           | ||| |
Db    1125 TGGAAAAC 1132


RESULT   5
V21209_06/c
Continuation (7 of 17) of V21209 from base 600001 (Methanococcus jannaschii circular chr
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP      Fragment Name      Begin      End
WP      V21209_00              1    110000
WP      V21209_01         100001    210000
WP      V21209_02         200001    310000
WP      V21209_03         300001    410000
WP      V21209_04         400001    510000
WP      V21209_05         500001    610000
WP      V21209_06         600001    710000
WP      V21209_07         700001    810000
WP      V21209_08         800001    910000
WP      V21209_09         900001   1010000
WP      V21209_10        1000001   1110000
WP      V21209_11        1100001   1210000
WP      V21209_12        1200001   1310000
WP      V21209_13        1300001   1410000
WP      V21209_14        1400001   1510000
WP      V21209_15        1500001   1610000
WP      V21209_16        1600001   1664379


  Query Match             3.6%;  Score 35.2;  DB 1;  Length 110000;
  Best Local Similarity   55.8%;  Pred. No. 3.9;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     652 gctgaagactatatgattgtgtacttgaatggtgcaaccccaagaaggaggatgccaggg 711
           ||| ||   |||  | |  | ||    ||||     |    || |||| ||| | |   |
Db   62905 GCTAAATTATATGAGTTGATATATAAAAATGTTAAGAATGAAAAAAGGCGGAAGAGTTGT 62846

Qy     712 ctaggctggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttgaaa 771
           ||   | | |||||||  |  ||  |||  ||  |||| | ||||  |  |||||  | |
Db   62845 ATAAAATAGTTGAAGAATTTATAAAAGAAAATGAACAGAAGATTGAAAATAAATTTAAGA 62786


RESULT   6
X29771
ID   X29771 standard; DNA; 1700 BP.
AC   X29771;
DT   22-JUN-1999  (first entry)
DE   L.lactis HsdM gene #2.
KW   HsdR; HsdM; HsdS; restriction/modification; bacteriophage; resistance;
KW   lactic acid bacterium; ds.
OS   Lactococcus lactis
PN   FR2767831-A1.
PD   05-MAR-1999.
PF   02-SEP-1997; 010885.
PR   02-SEP-1997; FR-010885.
PA   (INPG ) INRA INST NAT RECH AGRONOMIQUE.
PI   Chopin MC, Clier F, Ehrlich SD, Gautier M, Schouler C,
DR   WPI; 99-183265/16.
DR   P-PSDB; Y04661.
PT   Bacteriophage resistance mechanism subunit polypeptides - of lactic
PT   acid bacteria, especially Lactococcus lactis
PS   Claim 5; Page 39-41; 65pp; French.
CC   The invention relates to polypeptides that constitute the HsdR, HsdM or
CC   HsdS subunit of a R/M [restriction/modification] type Ic bacteriophage
CC   resistance mechanism active against the phages of lactic acid bacteria,
CC   especially Lactococcus lactis.  This sequence represents the gene
CC   encoding the Lactococcus lactis HsdM subunit.  Nucleic acids encoding
CC   the HsdR, HsdM or HsdS fragment can be used for expression of at least
CC   one bacteriophage resistance mechanism in a lactic acid bacterium.
SQ   Sequence   1700 BP;    541 A;    299 C;    368 G;    492 T;

  Query Match             3.5%;  Score 34.2;  DB 1;  Length 1700;
  Best Local Similarity   54.3%;  Pred. No. 0.83;
  Matches   69;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     200 atatcaatctttctctggaccaaagtgaaggatctattctctctgatgataacttggaga 259
           |||  | | ||||| |  || |||    || |  ||||       ||| |    |  |
Db    1360 ATAAAATTGTTTCTATCCACAAAAAACGAGAAGATATTGAAAGATATGCTCATGTTGCAA 1419

Qy     260 gtccagatgaaaattgacatcaatgttgtgaacttgatatcccgatgaagagattctt 319
           ||   ||||| ||  |    ||||   | |  |  ||| |||   | | | |||| | |
Db    1420 GTTTTGATGAGATCCAAGAAAATGACTTTAACTTAAATATCCCTCGTTATGTAGATACCT 1479

Qy     320 ttgagta 326
           ||||| |
Db    1480 TTGAGGA 1486


RESULT   7
Q81792
ID   Q81792 standard; DNA; 8478 BP.
AC   Q81792;
DT   26-SEP-1995  (first entry)
DE   B. subtilis biotin operon and flanking sequences.
KW   Biotin operon; recombinant production; dietary additive;
KW   animal feed; vitamin supplement; research reagent; ds.
OS   Bacillus subtilis.
FH   Key             Location/Qualifiers
FT   promoter        324  .352
FT                   /*tag=  a
FT                   /note= "sigma A promoter"
FT   misc_feature    355. .387
FT                   /*tag=  b
FT                   /note= "potential regulatory site"
FT   cds             439. .1218
FT                   /*tag=  c
FT                   /product= bioW
FT   cds             1208  2554
FT                   /*tag=  d
FT                   /product= bioA
FT   cds             2544. .3713
FT                   /*tag=  e
FT                   /product= bioF
FT   cds             3710. .4405
FT                   /*tag=  f
FT                   /product= bioD
FT   cds             4408  .5415
FT                   /*tag=  g
FT                   /product= bioB
FT   cds             5484. .6671
FT                   /*tag=  h
FT                   /product= bioI
FT   cds             6748. .7509
FT                   /*tag=  i
FT                   /label= ORF 2
FT   cds             7695. .0
FT                   /*tag=  j
FT                   /label= ORF 3
FT   terminator      249. .291
FT                   /*tag=  k
FT                   /note= "rho-independent termination site"
FT   terminator      5423. .5462
FT                   /*tag=  l
FT                   /note= "rho-independent termination site"
FT   terminator      7501. .7543
FT                   /*tag=  m
FT                   /note= "rho-independent termination site"
PN   EP-635572-A.
PD   25-JAN-1995.
PF   13-JUN-1994; 108998.
PR   25-JUN-1993; US-084709.
PR   06-MAY-1994; US-239430.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Bower SG,  Perkins JB,  Pero JG,  Yocum RR;
```

```
DR   WPI; 95-053684/08.
PT   Biotin genes, and constructs derived from Bacillus subtilis - for
PT   improved production of recombinant biotin or biotin precursor for
PT   use in e.g. dietary supplements
PS   Example II; Fig 14; 75pp; English.
CC   Q81792 is the B. subtilis biotin operon and flanking sequences,
CC   as part of an expression vector it can be used in the recombinant
CC   production of biotin (or biotin precursor protein). The biotin can
CC   be used as a dietary additive in animal feeds, and as a vitamin
CC   supplement for human consumption. Biotin is also useful as a
CC   reagent for research, and diagnostic prodcedures.
SQ   Sequence   8478 BP;    2432 A;    1748 C;    2088 G;    2205 T;


  Query Match             3.4%;  Score 33.2,  DB 1,  Length 8478;
  Best Local Similarity   56.4%;  Pred. No. 4.1;
  Matches   62;  Conservative    0;  Mismatches   48;  Indels    0,  Gaps    0;

Qy     854 gtaaaattaaatatgtcaatagcttatcagaactcagtgggctgatcccaatggattgca 913
           | || ||||    ||| |       | || |  |||  | |||||| | || |
Db    6025 GCAATATTAATGGCTGTGCAGGCTATGGCATATTTCAAAGAGCTGATTCAAAAGAGAAAAC 6084

Qy     914 tccacattccagagagcatcatcaatattgacttgaagctgaaagaaaag 963
           |||| || | ||   || |||  ||   |||||| || |||||||
Db    6085 GCCACCCTCAACAGGATATGATCAGCATGCTCTTGAAGGGGAGAGAAAAG 6134


RESULT   8
V23857/c
ID   V23857 standard; DNA; 777 BP.
AC   V23857;
DT   31-JUL-1998  (first entry)
DE   Plant LAC enzyme DNA sequence.
KW   lignin biosynthetic pathway; eucalyptus; pine; transgenic plant;
KW   lignin content; tree processing; cellulose fibre; ss.
OS   Eucalyptus grandis.
PN   WO9811205-A2.
PD   19-MAR-1998.
PF   10-SEP-1997; NZ0112.
PR   11-SEP-1996; US-713000.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PI   Bloksberg LN, Grierson AW, Havukkala IJ;
DR   WPI; 98-207674/18.
PT   Sequences useful for modification of plant lignin content or
PT   structure - from Eucalyptus grandis (eucalyptus) and Pinus radiata
PT   (pine) are associated with lignin biosynthesis pathway, useful e.g
PT   in paper industry
PS   Claim 1; Page 43-44; 82pp; English.
CC   This sequence represents a fragment of the LAC enzyme coding sequence. It
CC   is an example of a DNA sequence of the invention, which are from
CC   Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) associated with
CC   the lignin biosynthesis pathway. Constructs containing the DNA sequences
CC   can be used to produce transgenic plants or plant cells, especially woody
CC   plants e.g. eucalyptus or pine species but also e.g. monocotyledons or
CC   dicotyledons; by stably incorporating the constructs into the plant
CC   genome. The lignin content or structure, or activity of a specific enzyme
CC   in the plant, can therefore be modulated. Reductions in lignin content or
CC   changes in composition are useful in tree processing for paper. High
CC   lignin content results in energy- and chemical-intensive separation
CC   methods in order to obtain the pure cellulose fibre required. Reductions
CC   in lignin content may also be useful for forage crops, whilst increases
CC   or changes in composition may be desirable to increase the mechanical
CC   strength of wood, change its colour or increase its resistance to rot.
CC   The sequences are also useful as probes to isolate DNA sequences encoding
CC   enzymes involved in the lignin biosynthesis pathway from other plant
CC   species.
SQ   Sequence   777 BP;    252 A;    167 C;    193 G;    165 T;


  Query Match             3.4%;  Score 33;  DB 1;  Length 777;
  Best Local Similarity   49.7%;  Pred. No. 1.3;
  Matches   84;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     260 gtccagatgaaaattgacatcaatctggatgaaatttgatagccctgatgaagcaaattctt 319
           |||| ||||       ||| |                                        .
Db     469 GTCCTCATGATCAAAATGATCAATATTTTGGAAGATATATTTGAATTGAAGATATTTTTT 410

Qy     320 ttgagtacactgcccatgatcctgatgtttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 379
           |||| | ||  |||| ||                                        .
Db     409 TTTAGGCCCCTTGCCATTGTCGAATAAGAATTGCATCTTTAGTTGAAATTTTTTTATC 350

Qy     380 ttccagaatatacggccgaaagatgtacctaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 439
           |||'|'|  | |  || | |                                        .
Db     349 TTCCAGATGGCAGTGCAGGAACTAAACTTCCTGTATTGTCTGCTTGAAAT 300


RESULT   9
V21209_00/c
Continuation (4 of 17) of V21209 from base 80,001 (Methanococcus jannaschii circula
WP Sequence split into 17 fragments  [CCDS V212 ] Accession V212.4
WP      Fragment Name          Begin      End
WP      V21209_00                    1   100000
WP      V21209_01               100001   200000
WP      V21209_02               200001   300000
WP      V21209_03               300001   400000
WP      V21209_04               400001   500000
WP      V21209_05               500001   600000
WP      V21209_06               600001   700000
WP      V21209_07               700001   800000
WP      V21209_08               800001   900000
WP      V21209_09               900001  1000000
WP      V21209_10              1000001  1100000
WP      V21209_11              1100001  1200000
WP      V21209_12              1200001  1300000
WP      V21209_13              1300001  1400000
WP      V21209_14              1400001  1500000
WP      V21209_15              1500001  1600000
WP      V21209_16              1600001  1664976


  Query Match             3.4%;  Score 33;  DB 1;  Length 11
  Best Local Similarity   60.7%;  Pred. No. 1.9;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     602 tcatgaaaaatcttttactatatttttagtttttagatttatttagtttaaaaaaaa 661
           |||| | |||| ||| |  | | |                              .
Db   78525 TAATTGGAAATTTTATAAAAGATTTATAAATAATAGTGGAGATTATTTAAATTAAAAAAAA 78466

Qy     662 atatgattgtgtacttgaatggtgcaacc 690
           |||| | | | |  |||                .
Db   78465 ATATTAGTATCTCATTTATCTCTTTAACC 78437


RESULT  10
I75616/c
ID   I75616 standard; DNA; 10592 BP.
AC   I75616;
DT   20-FEB-1998  (first entry)
DE   Canine herpesvirus genomic DNA nCDS1592.
KW   CHV; vaccine; vector; gene therapy; canid; dog; CIR6; CUS2;
KW   CPK; serine-threonine protein kinase US3; gD1; glycoprotein D;
KW   CgG; glycoprotein G; glycoprotein 1; IgI; CUSP6; CUS4;
KW   unique short region; us.
OS   Canine herpesvirus.
FH   Key             Location/Qualifiers
FT   CDS             complement (377..1921)
FT                   /*tag= a
FT                   /product= CIR6
FT                   /note= "CHV IR6 (CIR6) gene"
FT   CDS             complement (1981..2254)
FT                   /*tag= b
FT                   /product= CUS2
FT                   /note= "CHV US2 (CUS2) gene"
FT   CDS             2375..3577
FT                   /*tag= c
```

```
FT                   /product= CPK
FT                   /note= "CHV serine threonine protein kinase US3
FT                        (CPK) gene"
FT   CDS             3689. .4936
FT                   /*tag=  d
FT                   /product= CgG
FT                   /note= "CHV glycoprotein G (CgG) gene"
FT   CDS             5129. .6165
FT                   /*tag=  e
FT                   /product= CgD
FT                   /note= "CHV glycoprotein D (CgD) gene"
FT   CDS             6225. .7319
FT                   /*tag=  f
FT                   /product= CgI
FT                   /note= "CHV glycoprotein I (CgI) gene"
FT   CDS             7469. .9035
FT                   /*tag=  g
FT                   /product= CgE
FT                   /note= "CHV glycoprotein E (CgE) gene"
FT   CDS             9028. .9264
FT                   /*tag=  h
FT                   /product= CUS8.5
FT                   /note= "CHV unique short region 8.5 gene"
FT   CDS             9376. .9735
FT                   /*tag=  i
FT                   /product= CUS9
FT                   /note= "CHV unique short region 9 gene"
PN   WO9729772-A1.
PD   21-AUG-1997.
PF   14-FEB-1997; U04115.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RA,  Haanes EJ;
DR   WPI; 97-424758/39.
DR   P-PSDB; W22979-83 AND W22983-3002.
PT   Recombinant canine herpes virus and its genome - useful as vaccine
PT   to protect canids against infectious, metabolic or genetic diseases
PS   Claim 22; Page 164-170; 240pp; English.
CC   This DNA sequence comprises genomic DNA molecule nCUS10592 of
CC   canine herpesvirus (CHV).  It includes coding sequences for CIR6,
CC   CUS2, CPK and CgG proteins (see W22979-83) and for CgD, CgI, CgE,
CC   CUS8.5 and CUS9 proteins (see W22983-3002).  CHV nucleic acids,
CC   encoded proteins and antibodies can be used to protect canids
CC   against CHV infection.  Live recombinant CHV can be used as a
CC   vaccine vector to protect canids against infectious, metabolic or
CC   genetic diseases, providing a vigourous and long-lasting immunity
CC   The genome can accommodate several sections of heterologous DNA
CC   (for making a multivalent vaccine) without interfering with
CC   packaging to live virus.  Fragments of CHV nucleic acids may also
CC   be used as probes and primers, and to inhibit CHV infection.
SQ   Sequence   10592 BP;    3757 A;    1746 C;    1681 G;    3408 T;
```

```
  Query Match             3.4%;  Score 32.8;  DB 1;  Length 10592;
  Best Local Similarity   53.0%;  Pred. No. 6.2;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
           || | ||| |||| | || || || | | | ||||| | | || |   | | |
Db    4610 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 4551

Qy     798 gttcatcagaacaatccttgctgtgacagaac tttataagttcaaaattcagcagtaa 857
           ||| | |||      || | |     || ||  | |  | |||| |
Db    1560 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 1501

Qy     858 aattaaatatgt 869
           |||||||  | |
Db    4490 AATTAAAAGTCT 4479


RESULT  11
T75590/c
ID   T75590 standard; DNA; 5495 BP.
```

```
AC   T75590;
DT   20-FEB-1998  (first entry)
DE   Canine herpesvirus genomic DNA nCUS5494.
KW   CHV; vaccine; vector; gene therapy; canid; dog; CIR6; CUS2;
KW   CPK; serine-threonine protein kinase US3; CgG; glycoprotein G;
KW   CgD; glycoprotein D; ds.
OS   Canine herpesvirus.
FH   Key             Location/Qualifiers
FT   CDS             complement (379. .930)
FT                   /*tag=  a
FT                   /product= CIR6
FT                   /note= "CHV IR6 (CIR6) gene"
FT   CDS             complement (1039. .2284)
FT                   /*tag=  b
FT                   /product= CUS2
FT                   /note= "CHV US2 (CUS2) gene"
FT   CDS             2384   3586
FT                   /*tag=  c
FT                   /product= CPK
FT                   /note= "CHV serine-threonine protein kinase US3
FT                        (CPK) gene"
FT   CDS             3698   4945
FT                   /*tag=  d
FT                   /product= CgG
FT                   /note= "CHV glycoprotein G (CgG) gene"
FT   CDS             5137   5495
FT                   /*tag=  e
FT                   /product= CgD
FT                   /note= "CHV glycoprotein D (CgD) partial sequence"
PN   WO9729772-A1.
PD   21-AUG-1997.
PF   14-FEB-1997; U04115.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RA,  Haanes EJ;
DR   WPI; 97-424758/39.
DR   P-PSDB; W22979-83.
PT   Recombinant canine herpes virus and its genome - useful as a vaccine
PT   to protect canids against infectious, metabolic or genetic diseases
PS   Claim 22; Page 125-128; 240pp; English.
CC   This DNA sequence comprises genomic DNA molecule nCUS5494 of canine
CC   herpesvirus (CHV).  It includes coding sequences for CIR6, CUS2,
CC   CPK and CgG proteins (see W22979-83) and a portion of CgI (see
CC   W22983).  CHV nucleic acids, encoded proteins and antibodies can be
CC   used to protect canids against CHV infection.  Live recombinant CHV
CC   can be used as a vaccine vector to protect canids against
CC   infectious, metabolic or genetic diseases, providing a vigourous
CC   and long-lasting immunity.  The genome can accommodate several
CC   sections of heterologous DNA (for making a multivalent vaccine)
CC   without interfering with packaging to live virus.  Fragments of CHV
CC   nucleic acids may also be used as probes and primers, and to
CC   inhibit CHV infection.
SQ   Sequence   5495 BP;    1917 A;    794 C;    948 G;    1836 T;
```

```
  Query Match             3.4%;  Score 32.8;  DB 1;  Length 5495;
  Best Local Similarity   53.0%;  Pred. No. 4.3;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaatttgaaatcatt atcattgttcatccatctt 797
           || |  ||| ||||   | | | || || || | | | |||| | |
Db    4619 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 4560

Qy     798 gttcatcagaacaatccttgctgtgacagaac tttataagttcaaaattcagcagtaa 857
           ||| | |||      || | |     || ||  | |  | |||| |
Db    4559 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 4500

Qy     858 aattaaatatgt 869
           |||||||  | |
Db    4499 AATTAAAAGTCT 4488


RESULT  12
```

```
V04600
ID   V04600 standard; cDNA; 610 BP.
AC   V04600;
DT   02-JUL-1998  (first entry)
DE   Flea serine protease gene SEQ ID NO:132.
KW   Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
KW   immunoglobulin protease, larvae, host animal; ss.
OS   Siphonaptera.
FH   Key             Location/Qualifiers
FT   CDS             1. .559
FT                   /*tag= a
FT                   /product= "serine protease"
FT                   /transl_except= (pos:133. .135,aa:Xaa)
FT                   /note= "Xaa= unknown"
PN   WO9740058-A1.
PD   30-OCT-1997.
PF   24 APR-1997; 006121.
PR   04-APR-1997; US-042945.
PR   24-APR-1996; US-639075.
PR   15-NOV-1996; US-749699.
PA   (HESK-) HESKA CORP.
PI   Frank GR, Gaines PJ, Grieve PB, Hunter SW, Pushlow KE,
PI   Silver G, Stiegler GL;
DR   WPI; 98-076762/07.
DR   P-PSDB; W41975.
PT   New flea protease genes and proteins - used in vaccine compositions
PT   for the prophylaxis and treatment of flea infestation, especially in
PT   cats or dogs
PS   Claim 1; Page 262-263; 318pp; English.
CC   The present sequence encodes a novel flea serine protease. The
CC   protease, its mimetopes, antibodies (Ab) and inhibitors of the
CC   protein, as well as the DNA encoding the protein, may all be used in
CC   therapeutic compositions to reduce flea protease activity (especially
CC   immunoglobulin protease) and so reduce flea infestation, especially in
CC   cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
CC   fleas which comprises anti-protease antibodies produced by a host animal
CC   in response to administration of the protein. Therapeutic compositions
CC   may further comprise a compound that reduces haematophagous ectoparasite
CC   burden by a method other than by reducing flea immunoglobulin protease
CC   activity. The novel flea DNA encoding the protein can also used to
CC   produce recombinant protein, and fragments of it are used as probes and
CC   primers for identification and isolation of related sequences, also as
CC   antisense, triplex-forming agents and ribozymes for inhibition of the
CC   synthesis of the protein. Ab are also useful for screening expression
CC   libraries, to purify the protein and to target cytotoxins to fleas.
SQ   Sequence   610 BP;    196 A;    96 C;    135 G;    181 T;

  Query Match             3.4%;  Score 32.8;  DB 1;  Length 610;
  Best Local Similarity   55.2%;  Pred. No. 1.3;
  Matches   64;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy     252 cttggacagtccagatgaaattgacatcaatgtggatgaacttgataccccgatgaagc 311
           |||  ||    || | || |||  |    ||| ||  ||  ||     |  |||||||
Db     171 CTTGAGCAACAAAGTTCAACTTGTGAAACTTGTAGAACAAGGTGTAGATCTGGATGAAGG 230

Qy     312 agattcttttgagtacactggccatgatcccacagccaacaaagattctggccaag 367
           |   | | || |   ||||||    | | | | ||| | | ||||||  |  | ||
Db     231 AACCTTTCTTAATGCCACTGGATGGGGTATAACAGCGACCGAAGATTTGCCTCCAG 286


RESULT  13
V66900/c
ID   V66900 standard; DNA; 5495 BP.
AC   V66900;
DT   07-JAN-1999  (first entry)
DE   Canine herpes virus nucleic acid sequence nCUS-5495 #1.
KW   Canine herpes virus; CHV; recombinant canine herpes virus vector;
KW   genome; vaccination; dog; protozoan; helminth; ectoparasite; bacteria;
KW   virus infection; ss.
OS   Canine herpes virus.
PN   US5804197-A.
PD   08-SEP-1998.
PF   12-JUL-1996; 680726.
PR   12-JUL-1996; US-680726.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RS, Haanes EJ;
DR   WPI; 98-505590/43.
PI   Canine herpes virus nucleic acids - useful for producing recombinant
PT   canine herpes virus vectors
PS   Claim 6; Column 47-54; 103pp; English.
CC   The present sequence represents a canine herpes virus (CHV) nucleic
CC   acid sequence. The nucleic acid sequence can be used for DNA vaccination
CC   of dogs against CHV and also other infective agents such as protozoans,
CC   helminths, ectoparasites, bacteria and viruses. CHV can be formulated by
CC   incorporation of heterologous nucleic acid molecules as a single
CC   multivalent therapeutic composition against a variety of canine
CC   pathogens.
SQ   Sequence   5495 BP;    1917 A;    794 C;    948 G;    1836 T;

  Query Match             3.4%;  Score 32.8;  DB 1;  Length 5495;
  Best Local Similarity   53.0%;  Pred. No. 4.3;
  Matches   70;  Conservative    1;  Mismatches   62;  Indels    1;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaaatttgaaatgattcatcattttcatgcatttt 797
           |   ||| |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    4619 GAAGGGTGATAGACAAATTATAAAAAGTTCATAATTTTGATTAATTTTAAATGACTATG 4560

Qy     798 gttcatcagaacaatccttgctgtgacacgaactttataagtcagaaattcagcgatgaa 857
           |||  || |||     |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    4559 GTTGTTCTGAAATTGATTACCAAACTAGCAATCAATAAACTTTGAGAATTCGTAGTTA 4500

Qy     858 aattaaatatgt 869
           ||||||  ||
Db    4499 AATTAAAAGTCT 4488


RESULT  14
V66901
ID   V66901 standard; DNA; 5495 BP.
AC   V66901;
DT   07-JAN-1999  (first entry)
DE   Canine herpes virus nucleic acid sequence nCUS-5495 #2.
KW   Canine herpes virus; CHV; recombinant canine herpes virus vector;
KW   genome; vaccination; dog; protozoan; helminth; ectoparasite; bacteria;
KW   virus infection; ss.
OS   Canine herpes virus.
PN   US5804197-A.
PD   08-SEP-1998.
PF   12-JUL-1996; 680726.
PR   12-JUL-1996; US-680726.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RS, Haanes EJ;
DR   WPI; 98-505590/43.
PT   Canine herpes virus nucleic acids - useful for producing recombinant
PT   canine herpes virus vectors
PS   Claim 6; Column 53-58; 103pp; English.
CC   The present sequence represents a canine herpes virus (CHV) nucleic
CC   acid sequence. The nucleic acid sequence can be used for DNA vaccination
CC   of dogs against CHV and also other infective agents such as protozoans,
CC   helminths, ectoparasites, bacteria and viruses. CHV can be formulated by
CC   incorporation of heterologous nucleic acid molecules as a single
CC   multivalent therapeutic composition against a variety of canine
CC   pathogens.
SQ   Sequence   5495 BP;    1836 A;    948 C;    794 G;    1917 T;

  Query Match             3.4%;  Score 32.8;  DB 1;  Length 5495;
  Best Local Similarity   53.0%;  Pred. No. 4.3;
  Matches   70;  Conservative    1;  Mismatches   62;  Indels    1;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaaatttgaaatgattcatcattttcatgcatttt 797
           |   ||| |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
```

```
Db    877 GAAGGGTGATAGACAAATTATTAAAAAGTTCAAACTTTCATTAATTTTAAATGACCACG 936

Qy    798 gttcatcagaacaatccttgctgtgacacgaa-ttttataagttcaaaattcagcagtaa 857
          ||| || |||      || |   || |  |     || ||  | |  | ||||  |
Db    937 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 996

Qy    858 aattaaatatgt 869
          ||||||'' ''
Db    997 AATTAAAAGTCT 1008


RESULT  15
V66905/c
ID   V66905 standard; cDNA; 1248 BP.
AC   V66905;
DT   07-JAN-1999  (first entry)
DE   Canine herpes virus nucleic acid sequence nCgG-1248.
KW   Canine herpes virus; CHV; recombinant canine herpes virus vector;
KW   genome  vaccination; dog; protozoan; helminth; ectoparasite; bacteria;
KW   virus infection; ss.
OS   Canine herpes virus.
FH   Key             Location/Qualifiers
FT   CDS             1. .1248
FT                   /*tag=  a
PN   US5804197-A.
PD   08-SEP-1998.
PF   12-JUL-1996; 680726.
PR   12-JUL-1996; US-680726.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RS, Haanes EJ;
DR   WPI; 98-505590/43.
DR   P-PSDB; W72649.
PT   Canine herpes virus nucleic acids - useful for producing recombinant
PT   canine herpes virus vectors
PS   Claim 6; Column 71-74; 103pp; English.
CC   The present sequence represents a canine herpes virus (CHV) nucleic
CC   acid sequence. The nucleic acid sequence can be used for DNA vaccination
CC   of dogs against CHV and also other infective agents such as protozoans,
CC   helminths, ectoparasites, bacteria and viruses. CHV can be formulated by
CC   incorporation of heterologous nucleic acid molecules as a single
CC   multivalent therapeutic composition against a variety of canine
CC   pathogens.
SQ   Sequence  1248 BP;   429 A;   169 C;   205 G;   445 T;


  Query Match             3.4%;  Score 32.8;  DB 1;  Length 1248;
  Best Local Similarity   53.0%;  Pred. No. 1.9;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    738 gatgattgacagaacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
          || |  ||| |||| | |  || || ||  | | ||||| | | || |  | |  |
Db    922 GAAGGGTGATAGACAAATTATTAAAAAGTTCAAACTTTCATTAATTTTAAATGACCACG 863

Qy    798 gttcatcagaacaatccttgctgtgacacgaa-ttttataagttcaaaattcagcagtaa 857
          ||| || |||      || |   || |  |     || ||  | |  | ||||  |
Db    862 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 803

Qy    858 aattaaatatgt 869
          ||||||'' ''
Db    802 AATTAAAAGTCT 791


Search completed: February  8, 2000, 20:40:11
Job time: 1019 sec
```

```
GM of: US-09-174-937-5 to: N_Geneseq_36:*  dat_format : pfs

Date: Feb 8, 2000 11:42 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_16/prod/cgn2_pub/dna_app/app_122611_135108_13316/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=1.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.000 -XGAPOP=12.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=12.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -SIAR1=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=myt -MINLEN=0
 -MAXLEN=10000000 -USER=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-174-937-5
Query length: 322
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 103.050000



score_list.
Sequence                 Strd Orig    DScore   DScore Len  ! Documentation ...
N_Geneseq_36:Isolate...   +   882.5  1944.91  1.9e-95   1986  ! Apoptosis-regulating protein N
N_Geneseq_36:X04121       +   882.5  1944.91  1.9e-95   1986  ! Human Nip1 encoding cDNA. Pol
N_Geneseq_36:X99873       +   579.0  1228.16  9.8e-61   1227  ! Gastric cancer associated gene
N_Geneseq_36:X99875       +   579.0  1228.16  9.8e-61   916   ! Gastric cancer associated gene
N_Geneseq_36:V31819       +   103.5   282.66  0.0142    2655  ! Solanum tuberosum tuber H-type
N_Geneseq_36:N71064       +    95.5   169.20  0.0299    1908  ! Gene encoding Plasmidium viva
N_Geneseq_36:V32449       +    91.0   149.29  0.14      5267  ! Human receptor tyrosine kinase
N_Geneseq_36:V34255       +    91.0   159.81  0.2664    1842  ! Human secreted protein gene 45
N_Geneseq_36:Q06184       +    90.5   146.52  1.46      5091  ! The iga gene, encoding IgA1 p
N_Geneseq_36:X04444       +    90.5   163.51  0.1658    1236  ! Yeast immunophilin FKBP46 cDNA
N_Geneseq_36:Q85965       +    89.5   145.09  1.76      4782  ! Rat homologue of human Wilsn
N_Geneseq_36:Q94644       +    89.0   148.59  1.12      2722  ! Recombinant cold-resistant pyr
N_Geneseq_36:T12675       +    89.0   148.59  1.12      2722  ! E. histolytica pyruvate phosp
N_Geneseq_36:T58840_2     +    88.0   104.19  933.77    110000  ! Continuation (3 of 6) of T5
N_Geneseq_36:V49597       +    88.0   166.73  0.2359    940   ! Human Osteosarcinoma cell line
N_Geneseq_36:V49592       +    88.0   157.31  0.3672    1316  ! Human Osteosarcinoma cell li
N_Geneseq_36:Q91449       +    87.5   143.17  2.85      3905  ! Tomato Cf-9 gene. Isolated pla
N_Geneseq_36:T06309       +    87.5   146.13  1.54      3050  ! Tomato pathogen resistance gen
N_Geneseq_36:T47877       +    87.5   145.98  1.57      3089  ! Tomato pathogen resistance Cf
N_Geneseq_36:Q27234       +    87.0   131.27  11.35     9506  ! Encodes APC gene in familial a
N_Geneseq_36:Q72097       +    87.0   130.29  11.76     10437  ! Adenomatous polyposis coli ti
N_Geneseq_36:Q70008       +    87.0   131.27  11.35     9496  ! Adenomatous polyposis coli pr
N_Geneseq_36:I95538       +    87.0   131.27  11.35     9506  ! Human adenomatous Polyposis co
N_Geneseq_36:T96158       +    87.0   131.27  11.35     9506  ! Human adenomatous Polyposis co
N_Geneseq_36:T27020       +    86.5   145.15  1.75      2760  ! Bacterial phosphoenolpyruvate
N_Geneseq_36:V72940       +    86.5   137.27  4.30      5921  ! Rat Munc13-1 encoding cDNA. S
N_Geneseq_36:Q40567       +    85.5   151.73  0.7506    1330  ! Mouse IL-1. Mouse interleukin
N_Geneseq_36:T96051       +    85.5   142.67  2.42      2930  ! DNA for bovine brain protein p
N_Geneseq_36:V34109       +    85.0   151.64  0.7596    1224  ! Human secreted protein gene 1
N_Geneseq_36:X20254       +    85.0   114.61  87.71     26776  ! Borrelia burgdorferi polynuc
N_Geneseq_36:Q33061       +    84.0   151.44  0.7798    1137  ! Plasmodium vivax circumsporozo
N_Geneseq_36:X23933       +    84.5   143.94  2.04      2124  ! N. meningitidis strain 88107
N_Geneseq_36:V74517       +    84.0   111.25  195.04    29555  ! Staphylococcus aureus contig
N_Geneseq_36:X13387       +    84.0   134.55  4.30      4234  ! Enterococcus faecalis genome
N_Geneseq_36:T06303       +    83.5   134.45  6.89      3905  ! Tomato pathogen resistance gen
N_Geneseq_36:V49600       +    83.0   133.13  8.16      3481  ! Murine protein tyrosine kinase
N_Geneseq_36:V47653       +    83.0   133.13  8.16      3981  ! Mouse protein tyrosine kinase
N_Geneseq_36:V21279       +    82.0   127.35  17.12     5979  ! Arabidopsis thaliana rRNA gene
N_Geneseq_36:V56447       +    82.0   121.90  31.89     8032  ! Human APC cDNA. Adenomatous po
N_Geneseq_36:V08391       +    82.0   139.15  3.77      2010  ! Human Parathyrtid hormone rece
N_Geneseq_36:V23264       +    81.5   132.75  8.77      3130  ! Nucleotide sequence of human S
N_Geneseq_36:X20248_01    +    81.5   [illegible]  [illegible]  ! Continuation (2 of 1
N_Geneseq_36:X20248_07    +    81.5   [illegible]  [illegible]  ! Continuation (8 of 1
N_Geneseq_36:X23322       +    81.5   134.64  5.26      [illegible]  ! N. meningitidis strain M
N_Geneseq_36:X39732       +    81.5   [illegible]  [illegible]  ! Bacillus pertussis associated

seq_name: N_Geneseq_36:T39228

seq_documentation_block:
ID   T39228 standard; cDNA; 1386 BP.
AC   T39228;
DT   18-DEC-1996 (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA.
KW   Apoptosis-regulating protein; Nip2 [illegible] BIB 1-K protein;
KW   Bcl-2; oncogene; cell death; [illegible]
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214..1158
FT                   /*tag= a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998.
PT   Nucleic acids encoding apoptosis regulating protein [illegible]
PT   diagnosing and treating [illegible]
PS   Claim 1; Page 39-41; 60pp; English.
CC   A cDNA clones (T39227-39) [illegible]
CC   regulating proteins Nip1, Nip2, Nip3, Bip1A, Bip1B and Bip1C
CC   (Wilson-Weigel, Nip [illegible]
CC   libraries using a yeast two-hybrid strategy with adenovirus E1B
CC   19K protein (W01010) as bait. Bip [illegible]
CC   E1B-interacting (Wilson, as [illegible] E1B 19K and Bcl-2 [illegible]
CC   cells against apoptosis. Nip1, Nip2, Nip3 bind to [illegible]
CC   Bip1C bind to 19K and Bcl-2, and Bip1B binds to Bcl-2 [illegible]
CC   acids can be used in diagnostic [illegible]
CC   apoptosis-regulating proteins, or to regulate cell [illegible]
CC   Increasing or suppressing expression of an apop [illegible]
CC   such as bcl-2.
SQ   Sequence  1386 BP;  428 A;  249 C;  [illegible]



alignment_scores:
                 Quality:  802.50          Length:     1
                   Ratio:  3.549           Gaps:       0
  Percent Similarity:  80.731   Percent Identity:  54.472

alignment_block:
 US-09-174-937-5 x T39228  ..

 Align seg 1,1  to  1-1144  from:  1  to:  1144

       22 GluAspValGlyMetAspIlePr[illegible]lyVal[illegible]hr 54
          |  |||  |||  |||||||||||  |||  |||||||||||  |||  |||
      250 GAAGATTTTGGAATAGATTCA[illegible]AAT A[illegible]AA A TAAA 2

       55 hAlaAlaAspMetArgProGl[illegible]sP hAspSerL ysyl yAsnA PP
             ||  ||||  |||||  |||  |||  |||||  |||  |||  |||  ||
      300 ACTAGCTATAACTGGACCAGAGCAT[illegible]TGCTCTAAAA TAAAT 34

       85 spThrHisProArgArgIlLys[illegible]sAspSerL ysAsnLysSerSer 71
             |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||
      351 GAAATAAAGTGAGAAAG[illegible]AAA AAAT T[illegible]AGCTCTA 37

       72 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 89
             |  |||  |||  |||  |||  |||  |||  |||  |||  |||  |
      397 CTTGATCCTAGTGATGGTTCTATTCTTTCTGATGATAAT CTTGATAGTCC 44

       90 sAspGluIleAspIleAsnValAr[illegible]sThrPr[illegible]A 1 4
             |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```



```
seq_name: N_Geneseq_36:X04120

seq_documentation_block
ID   X04120 standard; cDNA; 1386 BP.
AC   X04120;
DT   18-APR-1999  (first entry)
DE   Human Nip2 encoding cDNA.
KW   Human; Nip1; Nip2; Nip3; BipIA; Bip5; Bip13; adenovirus; cell death;
KW   viral infection; Bcl-2; proto-oncogene; mutational analysis; apoptosis;
KW   E1B 19K protein; cell survival regulation; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             214. .1158
FT                   /*tag= a
PN   US5878678-A
```

```
PD   12-JAN-1999.
PF   21-MAR-1995; 408095.
PR   02-AUG-1994; US-284139.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 99-152099/13.
DR   P-PSDB; W94324.
PT   Polypeptides that bind to anti-apoptotic proteins useful for
PT   protecting against cell death induced by viral infection and to
PT   modulate response to physical and chemical stimuli
PS   Claim 2; Column 51-54; 41pp; English.
CC   The present invention describes: (1) a method for regulating cell death,
CC   comprising exposing an isolated cell to a polypeptide selected from
CC   Nip1, Nip2, Nip3, BipIA, Bip5 and Bip13; (2) a method for neutralising
CC   the activity of the adenovirus E1B 19 kD protein, the Bcl-2 protein or
CC   the BHRF-1 protein, comprising exposing an isolated cell to a
CC   polypeptide as in (1); and (3) a method for detecting molecules that
CC   bind to at least one polypeptide as in (1), comprising lysing cells,
CC   exposing the lysate to the polypeptide and detecting any molecular
CC   polypeptide aggregates. The methods are useful for providing proteins
CC   able to bind to other proteins known to regulate cell survival e.g. it
CC   is known that E1B 19K protein provides a survival function similar to
CC   the cellular protooncogene bcl-2 gene product which is able to block
CC   apoptosis in haematopoietic B and T cells. The present sequence encodes
CC   human Nip2, given in the present invention.
SQ   Sequence   1386 BP;    428 A;    249 C;    386 G;    323 T;
```

```
alignment_scores:
           Quality:  862.50              Length:    801
             Ratio:  3.549                 Gaps:      6
 Percent Similarity:  80.731   Percent Identity:  58.472

alignment_block:
 US-09-174-937-5 x X04120   ..

 Align seg 1/1  to: X04120  from: 1  to: 1386
```



```
     638 IIGAIAIGAAGGCAAIIGAACCCIATAAAAAAAGIIAICAGCCAIGGG... 684

     171 AspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAl 187
         ||||||||||||||||||||||||||||:| ||||:::|||||||||::
     685 .....GGAIAIIAIGGGGAIGGAIIAAAIGGCAIIGIIGIAIIIGGIGI 728

     188 aCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluA 204
         :|||||:: :|:  :|||||  :|:  :  |||:::|||.:::|
     729 CGGIIICAIGCCIGAAASGIAGICAGCCIAACIIAIAGAIACCIGAIGGACA 778

     204 snLeuPheLeuTyrValIleSerThrLeuGluLeuMetValAlaGluAsp 220
         |||||||   |||||||||:::|||||| | |||:::|||||||||:::
     779 AICIIIIIAAAIAIGIIAIIGGCACIIIGGAGCIAIIAGIAGCAGAAAAC 828

     221 TyrMetIleValTyrLeuAsnGlyAlaThrProArgArgMetProGl 237
         ||||||||||||||||||||||||||||   ||||||:::|||||||::
     829 IACAIGAIAGIIIAIIIAAATGGTGCAACAACICGAAGAAAAATGCCCAG 878

     237 yLeuGlyIrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgL 254
         :|||||||||:::...|||||||||||||.:::|||||||||||||||||
     879 ICIGGGAIGGGIGAGGAAATGTTATCAGCAAAIIGATAGAAGGIIACGGA 928

     254 ysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThr 270
         |||||||||||||||...|||||||||||||||||:|||||||||||||||
     929 AAAAICIAAAATCCCIAATCAIIGIACAICCIICIIGGIIIAICAGAACA 978

     271 IleLeuAlaValThrArgProPheIleSerSerLysPheSerSerLysIl 287
         :::|||||||||||||||||||||||||||||||||||:|||||:..|||||
     979 CIICIGGCIGIIACAAGACCAIIIAIIAGCICGAAAIICAGCCAAAAAAAI 1028

     287 eLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspC 304
         |:.|||||  ...|||...|||||: ' |||.:||||||:::
    1029 IAGAIACGIGIIIAATIIGGCAGAACIAGCAGAACIIGICCCCAIGGAAI 1078

     304 ysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeuLysGlu 320
         ...  ||||||||  ||  :::...:|  :|||||:::
    1079 ACGIIGGCAIACCAGAAIGCAIAAAACAAGIIGAICAAGAACIIAAIGGA 1128

     321 Lys 321
         |||
    1129 AAA 1131

seq_name: N_Geneseq_36:X39873

seq_documentation_block:
ID   X39873 standard; DNA; 1227 BP.
AC   X39873;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen, diagnosis, research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; WO1467A.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 579-580; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   1227 BP;    388 A;    270 C;    334 G;    284 T;


alignment_scores:
            Quality:  579.00              Length:    1227
              Ratio:  4.559                 Gaps:       1
 Percent Similarity:  94.074   Percent Identity:  88.889

alignment_block:
 US-09-174-937-5 x X39873   ..

 Align seq 1/1  to. X39873  from. 1  to. 1227

       1 MetGluGluGluThrGluPheLeuGluLeuGlyThrArgIleSerArgPr 17
         ||||||||||||||||||||  |  |  |  |  |  |  |  |  |
     825 ATGGAGCACGAGACAGAGTTCCTTGACCTGGAACGCGAATATCAAGACC 874

      17 oAsnGlyLeuLeuSerGluAspValGlyMetAspIleProPheGlGlu 34
         |||||||||||||||||||||||  |  |  |  |  |  |  |  |
     875 AAATGGACTACTGTCAGACCATCTACGAATGGACATCGCTTTGAAGAGA 924

      34 lyValLeuSerProSerAlaAlaAspMetArgProGluProProAsnSer 5.
         |||||  |||||||||  |  |||  |  |  |  |  |  |  |
     925 GCGIGCIGAGTCCCAGIGCIGCAGACATGAGGCCIGAACCICCIAATTCI 974

      51 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaArgAs 67
         .||||||  |||||  |||  |  |  |  |  |  |  |  |  |
     975 CTGGATCIIAATGACACICATCCICGGAGAATCAAGIICACAGCCICAAA 1024

      67 nIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspA 84
         |||||||||||||||||||  |||  |  |  |  |  |  |  |  |
    1025 IATCAATCIIICICIGGACCAAAGIGAAGGATCIIATICICIIIGAIGATA 1 74

      84 snLeuAspSerProAspGluIleAspIleAsnValAspGluLeuAspTh 1
         |  ||||||||||  |  |  |  |  |  |  |  |  |  |  |
    1075 ACIIIGGACAGICCAGATGAAAIIGACAICAACGIGGATGAACIIGAIAC 1124

     100 rProAspGluAlaAspSerPheGluTyrThrGlyHisAspProThr.... 115
         |||  |||||  |||  |||  |  |  |  |  |  |  |
    1125 CCCCGATGAAGCAGATICIIIIGAGIACCCIGGICCCAIGAACGAACCCAC 1174

     116 AlaAsnLysAspSerGlyGlnGluSerGluSerIle...ProGluTyrTh 131
         :::::::| |||  |||  |  |  |  |  |  |  |  |  |
    1175 ACCCACAAAGAIICIGCCCCAACGAAICGAGCCIIAIIICCNGGAGIATAC 1224

     131 r 131
         |
    1225 C 1225

seq_name: N_Geneseq_36:X39876

seq_documentation_block:
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
```

```
PD   28-JAN-1999
PF   15-JUL-1998; 2146?9.
PR   22-JUN-1999; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    209 G;    184 T;


alignment_scores:
            Quality:  579.00            Length:    135
              Ratio:  4.559              Gaps:      4
 Percent Similarity:  94.074   Percent Identity: 88.889

alignment_block:
 US-09-174-937-5 x X39876   ..

 Align seq 1/1  to: X39876  from: 1  to: 816

       1 MetGluGluGluThrGluPheLeuGluLeuGlyThrArgIleSerArgPr 17
         ||||||||||||||||||||||||||||||||||| |||||||||||||||
     414 ATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGAACCAGGATATCAAGACC 463

      17 oAsnGlyLeuLeuSerGluAspValGlyMetAspIleProPheGluGluG 34
         |||||||||||||||||||||||||||||||||| ||||||||||||||||
     464 AAATGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGG 513

      34 lyValLeuSerProSerAlaAlaAspMetArgProGluProProAsnSer 50
         |||||||||||||||||||||||||||||||| ||  ||||||||||||||||
     514 GCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCT 563

      51 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaProAs 67
         |||||||||||||||||||||||||||||| ||  ||||||||||||||||
     564 CTGGATCTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAA 613

      67 nIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspA 84
         |||||||||||||||||||||||||||||| ||  ||||||||||||||||
     614 TATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATA 663

      84 snLeuAspSerProAspGluIleAspIleAsnValAspGluLeuAspTh 100
         |  ||| ||| ||||||||||||||||||  ||  ||||||||||||||||
     664 ACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATAC 713

     100 rProAspGluAlaAspSerPheGluTyrThrGlyHisAspProThr... 115
         |||||||||||||||||||||||||||||  ||  |||:::  :::
     714 CCTCCGATGAAGCAGATTCTTTTGAGTACCCTGGTCATGAAGAATCCCAC 763

     116 AlaAsnLysAspSerGlyGlnGluSerGluSerIle...ProGluTyrTh 131
         ......||||||||||||||  ::|| |::: ||:::     '.'' '
     764 AGCCACAAAGATTCTGGCCCAAGAAGTCAGTCTATTTCN?GAATATAC 813

     131 r 131
         |
     814 C 814

seq_name: N_Geneseq_36:V32919

seq_documentation_block:
ID   V32919 standard; DNA; 2655 BP.
AC   V32919;
DT   09-NOV-1998  (first entry)
DE   Solanum tuberosum tuber H-type alpha glucan phosphorylase gene.
KW   H-type, tuber, alpha-glucan phosphorylase, cold storage, chips;
KW   cold-sweetening; ds.
OS   Solanum tuberosum.
FH   Key             Location/Qualifiers
FT   CDS             12..2528
FT                   /*tag= a
FT                   /product= alpha-glucan H-type tuber phosphorylase
FT   mat_peptide     12..2525
FT                   /*tag= b
PN   WO9835051-A1.
PD   13-AUG-1998.
PF   05-FEB-1998; CA0055.
PR   04-JUN-1997; US-868786.
PR   10-FEB-1997; US-036946.
PA   (MIAC ) CANADA DEPT AGRIC & AGRI FOOD CANADA.
PI   Armstrong JD, Kawchuk LM, Knowles NR, Lynch DR;
DR   WPI; 98-447248/38.
DR   P-PSDB; W49078.
PT   Transgenic potato plants with improved cold storage characteristics
PT   - are modified plants with reduced alpha glucan L or H type
PT   phosphorylase activity, and have reduced cold sweetening
PS   Claim 7, Page 54-58, 91pp; English.
CC   The sequence is that of the alpha-glucan H-type tuber phosphorylase
CC   gene which can be used in the production of a potato plant with
CC   improved tuber cold-storage characteristics. Such plants can be
CC   used in agricultural potato production. The potatoes are especially
CC   useful in the manufacture of potato chips. The method reduces
CC   cold-sweetening (conversion of starches to sugars), resulting
CC   in the ability to store potatoes cooler temperatures. This
CC   produces prolonged dormancy, reduced incidence of disease and
CC   increased storage life. The potatoes also have improved specific
CC   gravity and improved chip score.
SQ   Sequence   2655 BP;    779 A;    516 C;    614 G;    746 T;


alignment_scores:
            Quality:  103.50            Length:    301
              Ratio:  0.745              Gaps:      17
 Percent Similarity:  46.179   Percent Identity: 21.262

alignment_block:
 US-09-174-937-5 x V32919   ..

 Align seq 1/1  to: V32919  from: 1  to: 2655

      50 SerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaPr 66
         ........||||||||||||||||||||            |||-|'   ||
    1008 GCGATACAACTAAATGACACACATCCA.........ACTCTTACGATTCC 1048

      66 oAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspA 83
         |:::::          |||                  :::::::' :
    1049 AGAGCTGATGCGGTTGCTA.....  ...      ATGGATGATG 1.??

      83 spAsnLeuAspSerProAspGluIleAspIleAsnValAspGluLeuAsp 99
         .....|||          ::::  :::||:::
    1078 AAGGACTTGGGTGGGATGAATCTTGGAATATCACT ............... 1112

     100 ThrProAspGluAlaAspSerPheGlyTyrThrGlyHisAsp......Pr 114
                  :::   .:.: :    '' ||::: ||
```

```
 1113 .......     ACTAGGACAATTGCCTATACGAATCATACAGTCCTACC 1150

  114 oThrAlaAsnLysAspSerGlyGlnGluSer......GluSerIleProG 129
      |   |||   ...    ..:|||             ...        .|||:
 1151 TGAAGCACTTGAAAAATGGTCTCAGGCAGTCATGTGGAAGCTCCTTCCTA 1200

  129 luTyrThrAlaGluGluGluArgGluAspAsnArgLeuTrpMetThrVal 145
      ::.                 ||:::.  |||  .  ||:::      :||
 1201 GACATATGGAAAICATTGAAGAAAIIGACAAACGGTTTGTTGCTACAATA 1250

  146 ValIleGlyGluGluGluGlnArgIleAspMetLysValIleGluProTy 162
      ...          .  ...|||  |||  .  ..          |||
 1251 ATG....      .  TCAGAAAGACCTGATCTTGAGAATAAGATGCCTAG 1288

  162 rArgArgValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAspGlyL 179
      |:|:::::::::::|
 1289 CATGCGCATTTTGGATCAC    ..... . ..... .. .   1307

  179 euAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSerSerArg 195
                                          :... . : ...
 1308 ..          . ..........          ..AACGCCACAAAA 1319

  196 AlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyrValIleSerTh 212
          |||     |||   ||||||   :::           ||
 1320 CCTGTTGTGCAT.     ATGGCTAACTTGTGTGTTGTCTCTTCACATAC 1363

  212 rLeu..... .. .          GluLeuMetValAlaGlu.... 219
      |:::                   :::::::::   ||||||
 1364 GGTAAATGGTGTTGCCCAGCTGCATAGTGACATCCTGAAGGCTGAGTTAT 1413

  220 .....AspTyrMetIleValTyr  ..... . .......  .LeuAsn 227
           ||||||...   |||.::                    |||
 1414 TTGCTGATTATCTCTCTGTATGCCCCACCAAGTTCCAGAATAAGACCAAT 1463

  228 GlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLysCy 244
      |'|   ||||||||||||                 |||::::::   ||
 1464 GGTATAACTCCTCGTAGG...          .TGGATCCGATTTTG 1495

  244 sTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSerPheIleI 261
      |         |:.:::::    ...   |||   ||||||:::
 1496 TAGTCCTGAGCTGAGTCATATAATTACCAAGTGGTTAAAAACA  .  .. 1536

  261 leValHisProSerTrpPheIleArg   ThrIleLeuAlaValThrArg 276
          :::|||   :::    :::|||||     |||
 1539 .       GATCAATGGGTGACGAACCTCGAATTGCTTGCTAATCTTCGG 1580

  277 ProPheIle  ... .. .          .          SerSerLysPh 283
      |||                                   |||:::|||::
 1581 GAGTTTGCTGATAATTCGGAGCTCCATGCTGAATGGGAATCAGCCAAGAT 1630

  283 eSerSerLysIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuI 300
      :::....|||          ....  :.:  :::::::::|||   :
 1631 GGCCAACAAGCAGCGTTTGGCACAGTATATACTGCATGTGACAGGT...G 1677

  300 leProMetAspCysIleHisIleProGluSerIleIleAsnIleAspLeu 316
      ::  .:::'|          |||:.|||  .  .:|||    .
 1678 TGAGCATCGAT           CCAAATTCC'TTTTTGACATACAAGTC 1715

  317 Lys 317
      |||
 1716 AAA 1718

seq_name: N_Geneseq_36.N71064

seq_documentation_block:
ID   N71064 standard; DNA; 1908 BP.
AC   N71064;
DT   01-JAN-1980  (first entry)
DE   Gene encoding Plasmodium vivax sporozoite circumsporozoite protein.
KW   Immunogen; vaccine; malaria; immunodominant epitope; ss.
OS   Plasmodium vivax.
FH   Key             Location/Qualifiers
FT   CDS             187..1293
FT                   /*tag= a
FT                   /product= circumsporozoite protein
PN   WO8700533-A.
PD   29-JAN-1987.
PF   24-JUN-1986; 001373.
PR   12-JUL-1985; US-754645.
PA   (UYNY-) NEW YORK UNIV.
PA   (ARNO/) ARNOT D E.
PI   Arnot DE, Enea V, Nussenzweig RS, Nussenzweig V;
DR   WPI; 87-037250/05.
DR   P-PSDB; P70708.
PT   New Plasmodium vivax circumsporozoite protein - and synthetic
PT   peptide(s) containing its dominant epitope, useful in anti-malarial
PI   vaccines
PS   Disclosure; fig; 9; 32pp; English.
CC   The circumsporozoite protein of P. vivax sporozoites encoded by
CC   this gene is useful in the construction of an anti-malarial
CC   vaccine. The sequence consists of a central domain of P7 704
CC   repeated 19 times plus N- and C-terminal regions (partially
CC   homologous with the corresponding domains of P. cynomolgi and P.
CC   knowlesi. See also P70704-17, P70709 and N71065.
SQ   Sequence  1908 BP;    677 A;    353 C;    476 G;    418 T;


alignment_scores:
            Quality:   95.50              Length:    174
              Ratio:   1.194                Gaps:      7
 Percent Similarity:   51.948   Percent Identity: 28.571

alignment_block:
 US-09-174-937-5 x N71064   ..

 Align seq 1/1  to: N71064  from: 1  to: 1908

      10 LeuGlyThrArgIleSerArgPro.AsnGlyLeuLeuSerGlyAspVal 31
         ::::|          |||||   :::                        
     310 GTAGGACAAAGTGCTAGCCGAGGTAGAGGACTT.................33 344

      29 yMetAspIleProPheGluGluGlyVal..Ser.ProSerAla.AspM 47
         |   :::         |||||'|        :::   :::
     345 TGAGAACCCAGATGACGAGTAAGTACAT...AAAAAAAAAAAAAAATTGAA 394

      43 etArgProGluProProAsnSerLeAspLeuAsnAspThrHisProArg 59
         :::  |||'|   |          
     395 AGAAAGCAGAACCAAAAAAT.......................TCAGT 42

      60 ArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGluSerSe 79
         :::   |'|||       |'|:::                 ' :::::::
     421 GAAAATAAGCTGAAACAACCAGTA.............GATAAAGTAGA 455

      79 sGlySerIleLeuSerAsp.AspAsnLeuAspSerProAspGluIleAsp 98
         :|'|:::      :::|'| ::::::::: '       :::      :::
     456 TGGACAGCCAGCAGGAGACAGACTACAACCAGCAACCCCCCTCCTATAA 505

      93 .....IleAsnValAspGluLeAspThrProAspGlnAlaAspSerPh 110
          ||    ::::'|::::::   :::      :::::::::
     506 GAGATGGACAACCAGCAGGAGATAGAGGAGCTTGGACAACCAGGAATGAGAT 555

     107 eGluTyrThrGlyHisAspProThrAlaAsnLysAspSerGlyAsnSerS 124
         : |          :::      :::      :::          :
     556 AGAGCAGATGGACAGCCAGCAGATAGAGTAGATGGACAGCCATCCAGGAGA 605

     124 erGluSerIleProGluTyrThrAlaGluGluGluArgGlnAspAsnArg 141
         ::|'|::::::   :::                  :::       :::
     606 CAGAGCAGATGATGGACAACCACCAGCAGACAGCCCCCCTCCTACTAGTAG 655

     142 LeuTrpMetThrValValIleGlyGluGluGlnArgIleAspMetLys 157
                        |'||||                 ::::::::
     656 ..          .CAGGTGATA.........GATAGCTTCCACAAATCAGGA 694

     157 sValIleGlu 160
         :::'|||
     695 GGTGATAGAG 694
```

```
seq_name: N_Geneseq_36:V32449

seq_documentation_block:
ID   V32449 standard; cDNA; 5267 BP.
AC   V32449;
DT   12-OCT-1998  (first entry)
DE   Human receptor tyrosine kinase LMR1_h cDNA.
KW   Receptor tyrosine kinase; LMR1_h; human; signal transduction;
KW   cancer; neurodegenerative disorder; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             7..4138
FT                   /*tag= a
FT                   /transl_except= (pos:193..195, aa:Glu)
FT                   /transl_except= (pos:205..207, aa:Val)
FT                   /transl_except= (pos:214..216, aa:Phe)
FT                   /transl_except= (pos:217..219, aa:Ser)
FT                   /transl_except= (pos:220..222, aa:Glu)
FT                   /transl_except= (pos:235..237, aa:Ala)
FT                   /transl_except= (pos:238..240, aa:Thr)
FT                   /transl_except= (pos:241..243, aa:Val)
FT   misc_difference 400..427
FT                   /*tag= b
FT                   /note= "not included in translated sequence of Fig
FT                    7A"
PN   WO9822507-A2.
PD   28-MAY-1998.
PF   21-NOV-1997; 97US-0976526.
PR   22-NOV-1996; 96US-0031675P
PA   (SUGE-) SUGEN INC.
PI   Joho KE, Plowman GD;
DR   WPI; 98-312419/27.
DR   P-PSDB; W48842.
PT   New isolated receptor tyrosine kinase genes - which are expressed in
PT   neuronal tissues and tumour cells, useful as targets for
PT   neurodegenerative disorders or cancers
PS   Claim 1; Fig 6B-C; 87pp; English.
CC   This nucleic acid molecule (NAM) codes for LMR1_h (see W48842), a
CC   novel human receptor tyrosine kinase (RTK).  A partial clone was
CC   isolated from human cerebellum cDNA using primers (see V42006 and
CC   V42011; based on rat LMR1_r cDNA (see V32448).  A 5048 bp clone of
CC   LMR1_h was isolated from a lambda ZAP cDNA library constructed from
CC   SNB75 cell line RNA, using the partial clone as probe.  This clone
CC   spans all but the first 27 amino acids of the predicted human LMP1
CC   protein sequence.  NAMs (see V32449-58) coding for novel kinases
CC   designated LMR1, LMR2 and LMR3 (see W48841-49) have been obtained
CC   from rat, human and mouse sources.  Expression of LMR1 and LMR3 is
CC   highly restricted to neuronal tissues with minimal expression in
CC   other adult or embryonic organs or in human tumour cell lines.
CC   LMR2 expression is limited to adult neuronal tissues, but is also
CC   very abundantly expressed in other non-neuronal foetal tissues and
CC   in numerous tumour cell lines.  Based on restricted expression of
CC   all 3 LMRs to adult neuronal tissues and the up-regulation of LMP2
CC   in a wide variety of tumour cell lines, these proteins may be
CC   critical targets for neurodegenerative disorders or cancer.  Claimed
CC   probes and antibodies may be used for detecting neurodegenerative
CC   diseases or cancer.  NAMs encoding LMRs may be used in gene therapy
CC   or to make transgenic animals
SQ   Sequence  5267 BP;  872 A;  1767 C;  1743 G;  885 T;


alignment_scores:
            Quality:   91.50              Length:    228
              Ratio:  0.880                Gaps:     12
 Percent Similarity:  45.614   Percent Identity: 21.930

alignment_block:
 US-09-174-937-5 x V32449  ..

 Align seq 1/1  to. V32449  from. 1  to. 5267

      11 GlyThrArgIleSerArgProAsnGlyLeuLeuSerGluAspValGlyMe 27
```

```
    2384 GGACCCCGCCTGCCCCTTCCTTCCGTCCCCTCCCCATCCCAGCACCGCCC 2433

      27 tAspIleProPheGluGluGlyValLeuSerProSerAlaAlaAspMetA 44
         ..:|||  |||||  ...  ...||||.:|||  ||
    2434 CCCACTTCCCTCGGAGGAG...GCCAGTGCCCCCGACGCCTTGATGCCC 2480

      44 rgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArg 60
         |||:  .  |||...                ...
    2481 TGCCTGACTCTCCCACGCCTGCTAC[illegible]GCGAGGTG .....TCTGCC 2524

      61 IleLysLeu.....................ThrAlaProAsnIleAs 69
         |||||||||                    ......|||.......
    2525 ATCAAGCTGGCTTCTGCCCTGAATGCCAGCAGCAGCTCTCCCGAGGTGGA 2574

      69 nLeuSerLeuAspGlnSerGlu.              .    GlySerIleL 80
         .        .........|||                 ::::::::::
    2575 GGCAGCCAGCAGTGAGGATGAGGA[illegible]AGCCACTTCAGCCATCT 2624

      80 euSerAspAspAsnLeuAspSer          .      ProAspGluIle 91
         :::::||   ...   ||'::                  ||||||   :::
    2625 TCACCGACACGTCCAGCGACGGC[illegible]ACGAGCCCGGATCTCCTG 2674

      92 AspIle.... ..... .  .                    ...AsnValAs 96
                                                   :::::::'
    2675 CCAGCCTTCCGCTCTCTGCAGAAGCAGGTGGGGACCCCCGACTCCCTGGA 2724

      96 pGluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGlyHisA 113
         !:::||||||   |||'''''''|||::::::      |||      :
    2725 CTCCCTGGACATCCCGTCCTCAGCCAGTGATGGTGGCTATGAGGTCTTCA 2774

     113 spProThrAlaAsn.........................LysAspSer 120
         ::|||:::|||':::                        ||||||
    2775 GCCCGTCGGCCACTGGCCCCTCTCGA[illegible]GCCGGAGTGTGGACAGT 2824

     121 GlyGlnGluSerGluSerIle......ProGluTyrThrAlaGluGlu1 135
         |||   ::::::|||'''         ||||||'''''   :::.||
    2825 GGCTATGACACCGAGAACTATGAG[illegible]GAGTTTGTGCTCAAGGAGGC 2874

     135 nArgGlu..  ..........                ............. 137
         :::|||
    2875 GCAGGAAGGGTGTGAGCCCCAGG[illegible]ACTTGGCTTAAGAGTTT 2924

     138 .........     .AspAsnArgLeuTrpMetThrValValIleGlyGlu 149
                       '''''||||||      |||   :::   |':
    2925 AGGGCCCCGGGCCCGAGACACGGCTC    [illegible]CTCCTCAGTGGCTC 2971

     150 GlnGluGlnArgIleAspMetLysValIleGluProTyrArgArgValIl 166
         :::||'''                    :::||||||||||
    2972 AACGAGAAG  . .  .....       AATCCCTACCCA          2993

     166 eSerHisGlyGlyAspSerGlyTyrTyrGlyAsp 177
                  ||||||:::||:::::|||
    2993 .........  .GACTCTGCCTACTTCTCAGAC 3013
```

```
seq_name: N_Geneseq_36:V34256

seq_documentation_block:
ID   V34256 standard; DNA; 1842 BP.
AC   V34256;
DT   28-JAN-1999  (first entry)
DE   Human secreted protein gene 45 clone HTTDS54.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency, blood, allergy, renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity, osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens
PN   WO9839446-A2.
PD   11-SEP-1998.
```

PF 06-MAR-1998; W04492.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.
PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043315.
PR 11-APR-1997; US-043568.
PR 11-APR-1997; US-043569.
PR 11-APR-1997; US-043577.
PR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.
PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057761.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hi JS,
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR WPI; 98-509887/51.
DR P-PSDB; W75159.
PT New isolated human genes and the secreted polypeptides they encode
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 258-259; 447pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number and the clone ID is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. V34145), for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 70 novel genes and their fragments (nucleic acid
CC sequences: V34154-V34223). [illegible] which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 70 polynucleotides, based on
CC which tissues they are most highly expressed in. [illegible] described
CC uses;.
SQ Sequence 1642 BP; 383 A; 428 C; 419 G; 409 T;

alignment_scores:
Quality: 91.00 Length: 212
Ratio: 0.958 Gaps: 11
Percent Similarity: 44.811 Percent Identity: 29.11%

alignment_block:
US-09-174-937-5 x V34256 ..

Align seq 1/1 to: V34256 from: 1 to: 1642

113 AspProThrAlaAsnLysAspSerGlyGlnGlnSerGluSerIleProIle 162
232 GACAGTACAAGTCAAAAATGTAATGGATAAAGGT..... 267

129 uTyrThrAlaGluGluGluArgGlnAspAsnArgLeuTrpMetTrpValV 178
268 .....GGACGCCTATGGCTTTTATAAAT 294

146 alIleGlyGluGlnGluGlnArgIleAspMetLysValIleGluProTyr 195
290 AACTCTGTGAAAACCACAGG..... 309

163 ArgArgValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsnGlyL 212

```
                ::::::|||   |||/||:::::::: ||||||...
     310 .........CTGGGGCATCCTGGACATCAGAGT:GGCIATGGCICTCAAA 350

     179 eu......AsnAlaIleIleValPheAlaAlaCysPheLeuProAspSer 193
                 |||   ||||||:::||::::|||   ||||||
     351 CCCTGAGCAATGAGATCATCATGTTTGTGGCTGGCTTTTTGGAGGGTTAC 400

     194 SerArgAlaAspTyr.........HisTyrValMetGluAsnLeuPheLe 207
           :|||  :::     :::          |||||.:::      |||.||:::
     401 CICACTGCCCCACACAIGAAIGACCACIACACA......AACCTCTACCC 444

     207 uTyrValIleSerThrLeuGluLeuMet...ValAlaGluAspTyrMetI 223
           :::.||:::      :::::|||!      :.....|| ...|||
     445 ACAGCIGATCACGAAACCIICCATCAIGGATAAAGTGCAGGAIIITATG. 493

     223 leValTyrLeuAsnGlyAlaThrProArgArgMetProGlyLeuGly 239

     493 ................................................. 493

     240 TrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLe 256
                    :::      |||:::      ||||||||::
     494 ................. GAGAAGCAAGATAAGTGGACCCGGAAAAATAT 525

     256 uLysSerPhe............IleIleValHisProSerTrpPheIle. 268
           :|'|::::::          :::     |:     ::::::   :::
     526 CAAAGAATACAAGACTGATICAIIITGGAGACATACAGGCTATGTGATGG 575

     269 .......................          ArgThrIleLeuAla 273
                                          |||:::|'||||
     576 CACAAATAGATGGCCTCTATGTAGGAGCAAAGAAGAGGGCTATATTAGAA 625

     274 ValThrArgProPheIleSerSerLysPheSerSerLysIleLysTyrVa 290
           |||:::::::::          :..|||...
     626 GGGACAAAGCCAATGACC.............CI.:ICCAGAIICAGTICCI 663

     290 lAsnSerLeuSerGluLeuSerGlyLeuIleP.. 301
           :'||||||:::::::::|||      |||||||'
     664 GAATAGTGTTGGAGATCTATTGGATCTGATTCT' 697

seq_name. N_Geneseq_36.Q06164

seq_documentation_block:
ID   Q06164 standard; DNA; 5091 BP.
AC   Q06164;
DT   31-JAN-1991  (first entry)
DE   The iga gene  encoding IgA1 protease
KW   IgA1; vaccine; meningitis; gonnorhoea; allergies; ss.
OS   Haemophilus influenzae.
FH   Key             Location/Qualifiers
FT   cds             262. .4887
FT                   /*tag= a
FT                   /product=IgA1 protease
FT   -10_signal      235. .240
FT                   /*tag= b
FT   -35_signal      211. .215
FT                   /*tag= c
FT   rbs             248  253
FT                   /*tag= d
PN   WO9011367-A.
PD   04-OCT-1990.
PF   16-MAR-1990; DK0073.
PR   17-MAR-1989; DK-001308
PA   (KILI/) KILIAN M.
PI   Kilian M, Poulsen K;
DR   WPI; 90-320267/42.
DR   P-PSDB; P07304.
PT   Immunoglobulin A1 protease prodn. - by cloning from
PT   microorganisms for immunisation against immunoglobulin A1
PT   protease producing bacteria
PS   Disclosure; fig 2; 44pp; English.
CC   This iga gene is from H.influenzae serotype b strain HK368.  On
CC   transformation of E.coli cells with a vector contg. this gene,
CC   immunoglobulin (Ig)A1 protease is produced which is useful in a
CC   vaccine for e.g. meningococcal meningitis, gonorrhoea or allergic
CC   diseases.
SQ   Sequence   5091 BP;    1862 A;    873 C;    879 G;    1477 T;


alignment_scores:
           Quality:   90.50              Length:     317
             Ratio:  0.611                Gaps.       18
 Percent Similarity:  46.688   Percent Identity: 21.136

alignment_block:
 US-09-174-937-5 x Q06164   ..

 Align seq 1/1  to: Q06164  from: 1  to: 5091

       7 PheLeuGluLeuGlyThrArgIleSe:ArgProAsnGlyLeuLeuSerGl 28
         |||   :::::::|||         ::::||   :::   :::|||:::::
    1723 TTTGCTTCTGTAGGGATTGTAAGTGGTCGCTCAACTCTTGTGCTTAATAA 1772

      23 uAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSerA 40
         :|||     :::|||
    1773 TGATAAACAAGTAGAT      ........................... 1788

      40 laAlaAspMetArgProGluProProAsnSer..................  5
                                 |||||||||
    1789 ......................CCAAAITCAAIIIACTTTGGCTTTAGA 1815

      51 .......LeuAspLeuAsnAspThr                HisProAr 69
                |||||||||||||  :::                .  .
    1816 GGCGGTCGATTAGACTTAAACGGTAATTCACTAACCTTTGATCACATCAA 1865

      59 gArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerG 78
         :  :|||        |||     ...||    .....:   :::::::
    1866 AAATATTGATGATGGTGCAAGACTAGTTAATCATAATATGACTAATGCTT 1915

      76 luGly.................SerIleLeuSerAspAspAsnLeuAsp 96
         :::::                 |||:::::::::::|||   |||
    1916 CAAAIATAACGAITACTGGGGAAAGTTAAIIAIAAGATCTAAATACAATT 1965

      87 SerProAspGluIleAspIleAsnValAspGluLeuAspThrProAspGl 1 3
         :::|||   :::|||                        |||::::::::::!
    1966 ACTCCATATAATATA  .                     GACGCACCAGATGA 1994

     103 uAlaAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspS 122
         |   ...  ...  ...                              ||||| :
    1995 AGATAATCCTTATGCCTTT....        . CGACGGATTAAAGATG 2029

     120 erGlyGlnGluSerGluSerIleProGluTyrThr..............  137
          :|||||!          .....   :::' ||||
    2030 GAGGACAGCTCTATTTAAATTTGGAAAATTACACTTATTATGCGTTAAGA 2 79

     132 ......................  .AlaGluGluGluArgGluAspAs 149
                                 :::::::::    | |::::'
    2080 AAAGGTGCGAGCACTCGTTCAGAATTACCTAAAAATAGTGGCGAAAGCAA 2129

     139 nArgLeuTrpMetThrValValIleGlyGluGluGluGluArgIleAspM 158
         |:::   !!'''          :::|||:::   ::::::
    2130 TGAAAATTGGCTA......TATATGGGTAAAACTTCCGAT.......... 2168

     156 etLysValIleGluProTyrArgArgValIleSerHis............ 175
                |||      |||::::|::::::::|||
    2164  ....... . GAAGCCAAAAGAAAIGIAATGAAGCATATTAACAACGAG 2212

     169 ......GlyGlyAspSerGlyTyrTyrGlyAsp..............31 188
               :::| |  :::|||||:::::::::                ::
    2203 CGTATGAATGGCTTTAATGGTTATTTTGGCCAGCAAGAGGGTAAAAATAA 2252

     178 yLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSerSerA 197
         .  |||      .....:|||                        :::::::
    2253 CGGTAATCTAAATGTGACTTTT.....................AAAGGCA 2291

     195 rgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyrValIleSer 214
         .......:                ::::||   ||||||   :::   :::
```

```
    2282 AAAGTGAG..............CAAAATGGTTTTTATTAATAGGGGGT 2316

     212 Thr...LeuGluLeuMetValAlaGluAspTyrMetIleValTyrLeuAs 227
         |||   |||...                 ||||||||||
    2317 ACAAACGTTAATGGGGATTTAACGGTTGAAAAAGGCACGTTATTCCTTTC 2366

     227 nGly...AlaThrProArgArgArg   MetPr gGlyLeuGlyTrpMetL 242
         :|||      ||||||     |||     ||||||     |
    2367 AGGCAGACCAACACCGCACGTAAGAGATAIGGAGGTATTTCTTCGACAA 2416

     242 ysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSer 258
         |||||              |||   ||||||    ||:|||     :::
    2417 AAAAA. . . . . GATCGTCACTTTGCTGAAAAT    ...AAT 2445

     259 PheIleIleValHisProSerTrpPheIleArgThrIleLeuAlaValTh 275
         ::::::|||:.    .|||:::    |  :::::   |||:::||
    2446 GAAGTGGTAGTAGAAGATGACTGGATTAACGGCAATTTTAAAGCAACCAC 2495

     275 r 275
         |
    2496 A 2496

seq_name: N_Geneseq_36:X04444

seq_documentation_block:
ID   X04444 standard; cDNA; 1236 BP.
AC   X04444;
DT   27-APR-1999  (first entry)
DE   Yeast immunophilin FKBP46 cDNA.
KW   Immunophilin, moth, insect cell; nuclear; immunosuppression; drug;
KW   transplant, tissue graft, ds.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   CDS             1..1236
FT                   /*tag= a
FT                   /product= "immunophilin FKBP46"
PN   US5861498-A.
PD   19-JAN-1999.
PF   31-OCT-1996; 741134.
PR   01-NOV-1995; US-007163.
PR   31-OCT-1996; US-741134
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   Alnemri ES, Fernandes-Alnemri T, Litwack G;
DR   WPI; 99-130433/11
DR   P-PSDB; W68011.
PT   Isolated immunophilin FKBP46 nucleic acids - useful for developing
PT   products for the study and identification of immunosuppressive
PT   agents for treating e.g. transplant and tissue graft patients
PS   Disclosure; Column 25:38, 21pp; English.
CC   This sequence represents the cDNA encoding the immunophilin FKBP46 from
CC   the yeast Saccharomyces cerevisiae. FKBP46 is a novel insect cell nuclear
CC   immunophilin, which can be used to study and identify additional
CC   immunosuppressive drugs that bind to it. Such immunosuppressive agents
CC   can be used in treating transplant and tissue graft patients.
SQ   Sequence  1236 BP;  462 A;  214 C;  306 G;  254 T;


alignment_scores:
            Quality:   90.50            Length:    144
              Ratio:  1.257              Gaps:      5
 Percent Similarity:  [illegible]  Percent Identity: 26.389

alignment_block:
 US-09-174-937-5 x X04444

 Align seq 1/1  to: X04444  from: 1  to: 1236

      23 GluAspValGlyMetAspIleProPheGluGluGlyVal.....Le 36
         ||||||        |||      |||:::::::::|||       ||
     325 GAAGACGATGACGAGGACGATGAGTTCCAAGAATCCGTCCTTTTGACTTT 374

      36 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 53
         |||||||...|||              ...    . :|||||||||
     375 ATCTCCGGAAGCC.....................AAIAICAAIAATCTTTGGATT 417

      53 euAspAspThrHisProArgArgIleLeuThrAlaIleAsnIleAsn 66
         |||||   |||       ||||||||||||||   |||   |||
     418 TGACGATTACTCCACGAAGAAGAAGTTAATTCATCAGTGGT..... 454

      77 LeuSerLeuAspGlnSerIleIle                    86
                    |||   |||                   |||   |||
     454               .TGTTACGT A.T'TTTT GT GTTTTTT TTTA 499

      86 pSerProAspGluIleAspIleAsnValAspGlLeLeAspThrProAspS 115
         |||   |||       ||||||   |||                |||
     469 GCATCCATTTGATACTTCAATTTTAACGTCAACGAAGACGAAGAAGAAGAA 518

     103 luAlaAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLys... 138
         ||||||||   ||||||       |||  |          ||||||
     539 ACGCTGACATCTATGACAGTGAAAAAAGTAAGAACTGACGCCGACGTCGAAT 588

     119 ..............................AspSerGlyGlnGlySerGl 125
                                        |||   |||  | |||  |
     589 GAAATTATTGGCCACGACATGGACGACTTGGATGACGAAGAGGAAGAAATA 638

     125 uSerIleProGluTyrThrAlaGluGlnGlnArgGlnAspAsnArgLeuT 142
         |   |||   |||         |||||  |||
     639 AGTTCGTATTGAAGAAGTCCAAGAAGAAGAAGATGAAGAAGAAGATAATGAT.... 694

     142 rpMetThrValValIleGlyGluGlnGluGln 152
                       |||||  |||   |  |
     685               GGAGAAGAAGAAGAA 699

seq_name: N_Geneseq_36:Q85956

seq_documentation_block:
ID   Q85956 standard; DNA; 4782 BP.
AC   Q85956;
DT   14-FEB-1996  (first entry)
DE   Rat homologue of human Wilson disease gene ATP7B.
KW   Copper transport, ATPase; Wilson disease; [illegible]al markers; probe;
KW   yeast artificial chromosome library; hybridisation; Menkes disease;
KW   copper binding site, phosphate domain, transmembrane, transport;
KW   phosphorylation, ATP binding domain, autosomal recessive disorder;
KW   toxicity; liver; brain; gene therapy; heavy metal; rat; ss.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             1..4371
FT                   /*tag= a
FT                   /product= rat copper transporting ATPase
PN   WO9508641-A.
PD   30-MAR-1995.
PF   21-SEP-1994; CA0519.
PR   21-SEP-1993; CA-106602.
PR   21-OCT-1993; JA-108927.
PA   (HSCR-) HSC RES & DEV LP.
PI   Bull P, Cox DW, Thomas G;
DR   WPI; 95-119813/16.
DR   P-PSDB; R70419.
PT   Nucleotide sequence contg. the gene for Wilson disease - used to
PT   treat and detect Wilson disease
PS   Example 5; Fig 17; 127pp; English.
CC   The nucleotide sequence of the rat gene homologous to the human gene
CC   (Q85941, encoding the copper transporting ATPase (ATP7B), associated with
CC   Wilson disease (WD). The rat gene was isolated from a rat liver cDNA
CC   library with probes derived from the human WD gene sequence. Twenty-five
CC   clones were isolated which resulted in a consensus sequence of about
CC   4.7 kb. The rat gene encodes a single large open reading frame (ORF)
CC   with about 300 bp of 3' untranslated sequence. The rat and human gene
CC   share about 82% sequence homology at the amino acid level. The overall
CC   structure of the gene is similar to the human gene with the exception of
CC   the lack of copper binding domain 4 in the rat sequence. The rat gene
CC   disorder is found in Long-Evans Cinnamon rats which share many of the
CC   clinical and biochemical features with WD, so can be used as models for
CC   WD. The gene can be used to treat Wilson disease - an autosomal recessive
CC   disorder of copper transport which includes [illegible] and
```

```
CC   toxicity to the liver and brain.  The DNA sequence can also be used in
CC   gene therapy to remove heavy metals from an organ.
SQ   Sequence   4782 BP;    1159 A;    1299 C;    1283 G;    1041 T;


alignment_scores:
            Quality:   89.50              Length:    347
              Ratio:  0.617                Gaps:     13
 Percent Similarity:  41.787   Percent Identity: 17.867

alignment_block:
 US-09-174-937-5 x Q85956    .

 Align seq 1/1  to: Q85956  from: 1  to: 4782

      13 ArgIleSerArgProAsnGlyLeuLeuSerGluAspValGlyMetAspIl 29
      982 AAAGTATCCCTTCCCGATGGCCTAGAGAAGGAGAGTGGATCTTCCAGTGT 1031

       29 eProPheGluGluGlyValLeuSerProSerAlaAlaAspMetArgProG 46
     1032 CCCC            TCCCTTGGCTCCTCCCAGAGACAGCAGGAGCCAG 1069

       46 luProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLys 62
     1070 GCCCATGCAGGACTGCGGTACTCACCATCACTGGCATTCCCCGTGACTCG 1119

       63 LeuThrAlaPro                                .....As 67
     1120 TCTGTTCAGCCCATGGAAGACATGCTGTCCCAGATGAAGGGTGTGCAGCA 1169

       67 nIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspA 84
     1170 AATAGACACTCCTTTGGCAGAGGGAGTGGAGTAGTTCTTTACGATCCCT 1219

       84 snLeuAspSerProAspGluIleAspIleAsnValAspGluLeuAspThr 100
     1220 CAGTAGTTAGCTCGGATGAACTCCCGACGCCTCTAGAAGACATG...... 1263

      101 ProAspGluAlaAspSerPheGluTyrThrGlyHisAspProThrAlaAs 117
     1264                GGCTTTCAGGTGTCAGTCAATCCCGAAAACATTAC 1298

      117 nLysAspSerGlyGlnGluSerGluSerIleProGluTyrThrAlaGluG 134
     1299 TACTAACCGAGTCAGCTCTGGGAATTCTGTGCCACAA ............ 1335

      134 luGluArgGluAspAsnArgLeuTrpMetThrValValIleGlyGluGln 150
     1336 ...                                GCCGTGGGTGATTCA 1350

      151 GluGlnArgIleAspMetLysValIleGluProTyrArgArgValIleSe 167
     1351 CCAGGGTCTGTG...CAAAATATGGCTTCTGACACTAGAGGACTCCTCAC 1397

      167 rHisGlyGlyAspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleV 184
     1398 ACACCAAGGC   CCTGGCTACTTGTCAGACAGC ............ 1428

      184 alPheAlaAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyr 200
     1429                CCACCATCCCCTCTAGGA.... ... ... 1446

      201 ValMetGluAsnLeuPheLeuTyrValIleSerThrLeuGluLeuMetVa 217
     1447 .......                          ............AC 1448

      217 lAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgA 234
     1449 AGCATCACAGAAGTGCTTTGTACAGATCAAAGGCATGACC.. ..... 1488

      234 rgMetProGlyLeuGlyTrpMetLysLysCysTyrGlnMetIleAspArg 250

     1489 ..... TGTCCGTCCTGTGTGTCTAACATAGAAAGG 1518

      251 ArgLeuArgLysAsnLeuLysSerPheIleIleValHisProSerTrpPh 267
     1519 AGTCTGCAGAGACAT.....  . ..... 1533

      267 eIleArgThrIleLeuAlaValThrArgProPheIleSerSerLys.... 282
     1534 ....GCCGGTATTCTCTCCGTGTTCGTCGCCTTCATGTCGGGAAAGCCAG 1579

      282 ..... 282
     1580 AGGTCAAGTATGACCCAGAGGTCATCCAGTCTCCCAGGATAGCTCAGCTC 1629

      283 ..  .....PheSerSerLysIleLysTyrValAsnSerLeuSe 294
     1630 ATCGAGGACCTGGGCTTCGAAGCAGCAATCATGGAGCAGACACAGTCTC 1679

      294 rGlu..      .....LeuSerGlyLeuIleProMetAspCysI 305
     1680 TGAAGGTGACATCGAACTGATTATCACAGGCATGACCTGGCCTTCTTGTC 1729

      305 leHisIleProGluSerIleIleAsnIleAspLeuLysLeu 318
     1730 TTCAC. ..... .AACATAGAATCTAAGCTC 1752

seq_name: N_Geneseq_36:Q94644

seq_documentation_block:
ID   Q94644 standard; cDNA; 2722 BP.
AC   Q94644;
DT   25-JAN-1996  (first entry)
DE   Recombinant cold-resistant pyruvate phosphate dikinase cDNA.
KW   Recombinant pyruvate phosphate dikinase; cold-resistant;
KW   transformed plants; maize; deccan grass; millet; sorghum; ss.
OS   Flaveria sp.
FH   Key             Location/Qualifiers
FT   cds             48..2705
FT                   /*tag=  a
PN   WO9515385-A1.
PD   08-JUN-1995.
PF   01-DEC-1994; J02022.
PR   03-DEC-1993; AU-052275.
PR   29-JUL-1994; JP-197780
PA   (NISB ) JAPAN TOBACCO INC.
PI   Burnell NJ,  Ohta S,  Usami S;
DR   WPI; 95-215265/28.
DR   P-PSDB; R76710.
PT   Recombinant cold-resistant pyruvate phosphate di kinase - and DNA
PT   coding for it which can be used to impart cold resistance to plants.
PS   Claim 1; Pages 41-47; 70pp; Japanese.
CC   Q94644 encodes P76710 a recombinant cold-resistant pyruvate phosphate
CC   dikinase (PPDK). The recombinant PPDK can be used in the prodn. of
CC   transformed cold-resistant plants, e.g. maize, deccan grass, millet
CC   or sorghum.
SQ   Sequence   2722 BP;    969 A;    403 C;    558 G;    792 T;


alignment_scores:
            Quality:   88.00              Length:    374
              Ratio:  0.515                Gaps:     19
 Percent Similarity:  45.722   Percent Identity: 21.390

alignment_block:
 US-09-174-937-5 x Q94644   ..

 Align seq 1/1  to: Q94644  from: 1  to: 2722

       3 GluGluThrGluPheLeuGluLeuGlyThrArgIleSerArgProAsnGl 19
    1032 GAAAGAAAGAAACTTTACATGCTCCAAACTAGAAATGGAAAGATGAATGC 1081

      19 yLeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyValL 36
```

```
     :      :::      :::|||:::|||:::      ||||||||||::::
1082 AACTGCTACTGTCAGAACAGGAGTTGATATG...GTTGAAGAAGGACTTA 1128

  36 euSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAsp 52
     :..          :::  '|||||          |||:::|||:::|||
1129 TTACAAAAGAACAAGCCATTATGAGA...ATTGCACCACAATCAGTTGAT 1175

  53 .....LeuAsnAspThrHisProArgArgIleLysLeuThrAlaProAs 67
       |||:::   :::   |||   ...          |||||
1176 CAATTACTTCATAAGAATATGCCAGCTAATTAT   GCAGAAGCTCCATT 1222

  67 nIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspA 84
     :.....   ..:|||        |||   |||....   :::
1223 AGTTAAA...GGACTTCCAGCATCACCAGGACGTGCTACAGGAGCTGTTG 1269

  84 snLeuAspSerProAspGluIleAsp          ..... ....Ile 93
     .. |||       |||   .                      ...
1270 TTTTTGATGCCCATGATGCAGTTGAACAAGCTAAAGGAAAGAAAGTTGTT 1319

  94 AsnValAspGluLeuAspThrProAspGluAlaAspSerPheGluTyrTh 110
     .    :::          |||.....          ...|||
1320 CTTCTTAGAGAAGAAACTAAACCAGAAGATATTCATGGATTCTTTGTT.. 1367

 110 rGlyHisAspProThrAlaAsnLysAspSerGlyGlnGluSerGluSerI 127
                                                ...|||...|
1368 .   .   .   .   .   .                .....GCTGAAGGTA 1377

 127 leProGluTyrThr.AlaGluGluGluArgGluAspAsnArgLeu..... 141
     ||         ||| |||||||||||||:::::::   ::::::|||
1378 TT......TTAACCTGCAGAGGAGGAAAAACATCACACGCAGCTGTCGTT 1421

 142 ....      .TrpMetThrValValIleGlyGluGlnGluGln..ArgIl 154
               |||:....:   |||...  |||   :....  :....
1422 GCTAGAGGTATGGGTAAACCATGTGTTTCAGGAGCTGAAGGAATTAAAGT 1471

 154 eAspMet   lysValIleGluProTyrArgArgValIleSerHisGlyG 170
     :|'|:::   |||   |||                ::      |||   |
1472 TGATGTTGCTAAGAAAATTGCTAAGATTGGAAGCTTTGAAGTTCATGAAG 1521

 170 lyAsp          .   .   .              .... ..... 171
     ||||
1522 GAGATATTTTAACTATTGATGGATCAACTGGATGTGTCTATAAGGGAGAA 1571

 172 .....  .   .      .....SerGlyTyrTyrGly ..... 176
                           |||||| ||:::|||
1572 GTTCCATTAGAAGAACTACAAGTTGGATCAGTACATTTCGGAACCATCTT 1621

 177 ..        AspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheL 190
               :::   :::::   |||   |||||||||||      |
1622 AAAATGGGCCAATGAAATTAAAAAGATTGGACTTTTTGCTGCTGGAGATC 1671

 190 euProAspSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPhe 206
     |||||  ..          .   .              '||:  ....
1672 TTCCATCAGCTGCTAAGAAAGCCCTTGAATTTGGAGCTGAAGGTATTGGA 1721

 207 leu.  .         TyrValIleSerThrLeuGluLeuMetValAlaGl 219
     |||          :....  :::   |||   ...
1722 GTTTGTAGAATTGAATGTATGTTCAATGCAGTTGAAAGACTTCCAATTGT 1771

 219 sAspTyrMetIleVal .          .  .  .............TyrLeuA 227
           ||||||...                              |||||||
1772 TGTCAAGATGATTCTTTCAAATACCCTTGAAGAAAGAAAGAAATATCTTA 1821

 227 snGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLys 243
     ||      |||   :.......     ...|||   :::|||
1822 ATGAACTTATGCCACTTCAAAAACAAGATTCCATTGGATTATTGGAAGACT 1871

 244 ..................CysTyrGlnMetIleAspArgArgLeuArgLy 254
                      :.........|||          |||.....
1872 ATGAATGGACTTCCAGTCACTGTCAGACTTCTTGATCCACCATTACATGA 1921

 254 sAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThrI 271
     :                                    |||:::   |||:

     1922 A.....TTCCTTGAACTT 1942

      271 leLeuAlaValThrArgProPheIleSerLysPheSerGlys... 296
          ::      :::   |||   :::::::::      :::   :::
     1936 TTGAAGAGTTAATGAGAGAAATCTTCGAAATCGAAACTTCAGTTAACAGT 1985

      297 .........................IleLysTyrValAsnGelLeuGe 314
                                   ::: |      :::::
     1986 GAAGGACTTGCAGAAAAAGAAGTTGTTCTTAATAAAATTAAAGAACTTAT 2035

      294 rGluLeuSerGlyLeuIle 300
          '|:::::::   :::|'|
     2036 GGAAGTTAATCCAATGATT 2054

seq_name: N_Geneseq_36:T12675

seq_documentation_block:
ID   T12675 standard; DNA; 2722 BP.
AC   T12675;
DT   26-SEP-1996  (first entry)
DE   E. histolytica pyruvate phosphate dikinase coding sequence.
KW   Wild type; pyruvate phosphate dikinase; PPDK; cold-resistant;
KW   F. brownii; F. bidentis; maize; cold resistance; ss.
OS   Entamoeba histolytica.
FH   Key             Location/Qualifiers
FT   cds             48   2705
FT                   /*tag=  a
FT                   /product= PPDK
PN   WO9604369-A1.
PD   15-FEB-1996.
PF   30-MAY-1995; J01040.
PR   29-JUL-1994; JP-197780.
PR   01-DEC-1994; WO-JP02022.
PA   (NISB ) JAPAN TOBACCO INC.
PI   Burnell NJ,  Ohta S,  Usami S;
DR   WPI; 96-129896/19.
DR   P-PSDB; R90926.
PT   Recombinant cold-resistant pyruvate phosphate dikinase - imparts
PT   improved cold tolerance to transgenic plants which express the
PT   enzyme
PS   Claim 1; Page 41-47; 70pp; Japanese.
CC   The sequences given in T12672-75 encode wild type pyruvate phosphate
CC   dikinase (PPDK) enzymes from various organisms. These sequences
CC   were used in the design of a novel protein having cold-resistant
CC   PPDK activity (see also R90927). The cold-resistant PPDK is based
CC   primarily on the PPDP sequence of P. brittonii and comprises an
CC   addition, deletion or substitution of one or more amino acids in the
CC   C-terminal portion, esp. within the region residues 832-855. The
CC   cold-resistant PPDK may also be based on the F. bidentis PPDK
CC   sequence having Pro at position 850 and Ile, Val or Met at 852.
CC   Plants, e.g. maize, transformed with DNA encoding the cold-resistant
CC   PPDK, have improved cold resistance.
SQ   Sequence   2722 BP;    909 A;    403 C;    558 G;    792 T;

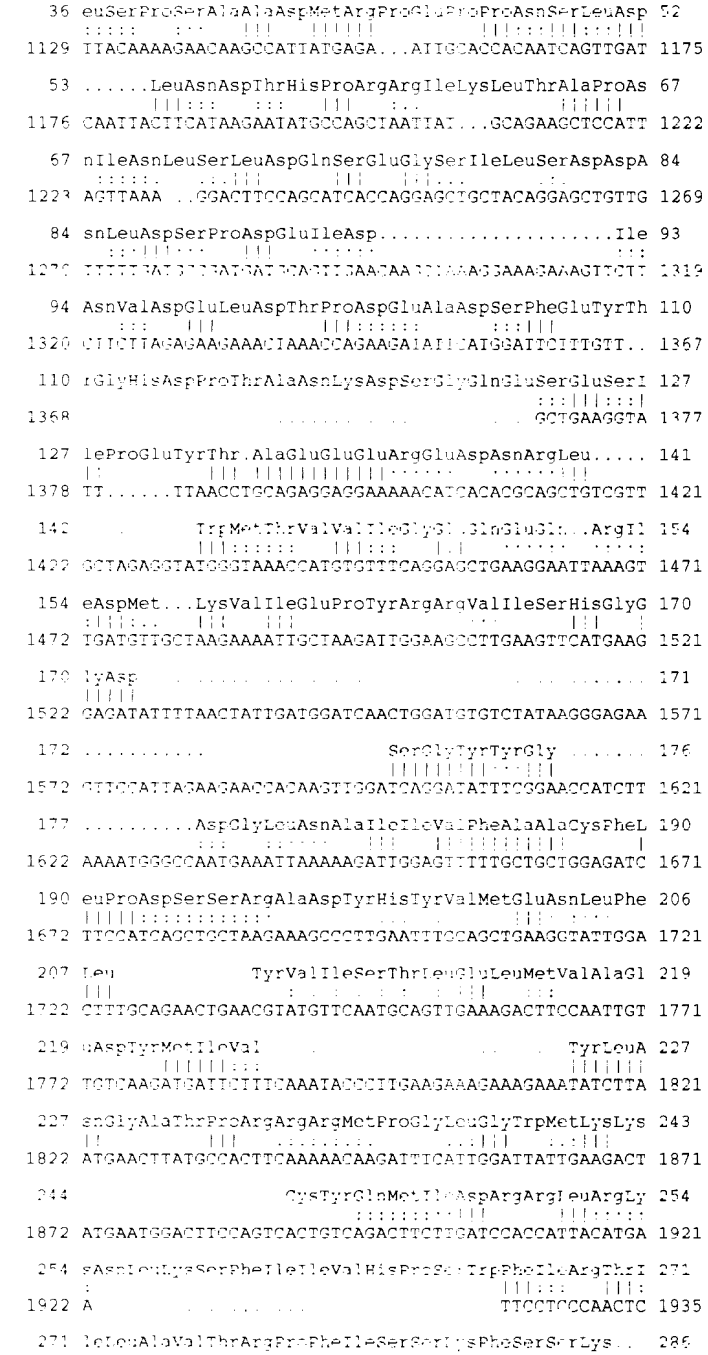

alignment_scores:
            Quality:   88.00              Length:     374
              Ratio:   0.519                Gaps:      19
 Percent Similarity:  45.722   Percent Identity: 21.842

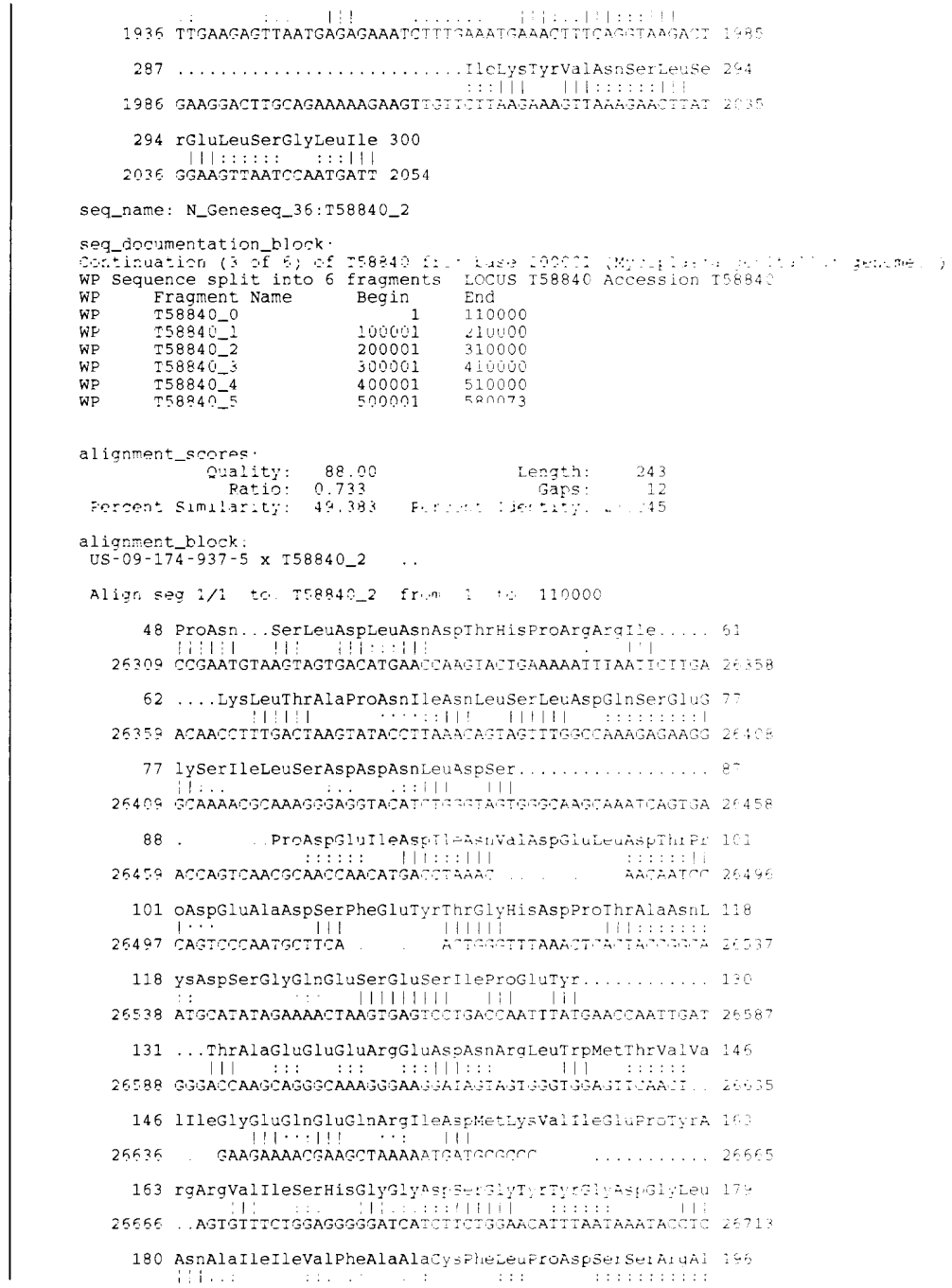
alignment_block:
 US-09-174-937-5 x T12675   ..

 Align seq 1/1  to: T12675  from: 1  to: 2722

       3 GluGluThrGluPheLeuGluLeuGluThrArgIleArgProAsnGl 19
         '|.:::   ::::::     |      '''    ::::::   .:::
    1032 GAAAGAAAGAAACTTTACATGCTCCAAACTAGAAATCCTAAGTTTAAGC 1081

      19 yLeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyVal 38
         :      :::      :::|'|:::.:' :::                ::::
    1082 AACTGCTACTGTCAGAACAGGAGTTGATATG   GTTGAAGAAGGACTTA 1128
```





```
seq_name: N_Geneseq_36:T58840_2

seq_documentation_block:
Continuation (3 of 6) of T58840 from base 100001 (Mycoplasma pulmonis genome.)
WP Sequence split into 6 fragments  LOCUS T58840 Accession T58840
WP      Fragment Name       Begin      End
WP      T58840_0                 1    110000
WP      T58840_1            100001    210000
WP      T58840_2            200001    310000
WP      T58840_3            300001    410000
WP      T58840_4            400001    510000
WP      T58840_5            500001    580073


alignment_scores:
            Quality:   88.00              Length:     243
              Ratio:  0.733                Gaps:      12
 Percent Similarity:  49.383   Percent Identity:  ...45

alignment_block:
 US-09-174-937-5 x T58840_2   ..

 Align seg 1/1  to  T58840_2  from  1  to  110000
```



```
seq_name: N_Geneseq_36:V49593

seq_documentation_block.
ID   V49593 standard; cDNA to mRNA; 990 BP.
AC   V49593;
DT   21-OCT-1998  (first entry)
DE   Human Osteosarcinoma cell line U-2 OS clone HP10304 cDNA #2.
KW   Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW   differentiation; immune system; stimulator; suppressor; regulator;
KW   haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW   haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS   Homo sapiens.
PN   WO9921328-A2.
PD   22-MAY-1998.
PF   07-NOV-1997; J04056.
PR   13-NOV-1996; JP-301429
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR   WPI; 98-297232/26
DR   P-PSDB; W64555.
PT   Human protein having transmembrane domain - useful for, e.g.
PT   research and nutrition
PS   Claim 3; Page 131; 205pp; English.
CC   V49550-V49599 are cDNA sequences which encode human proteins containing a
CC   transmembrane domain. These proteins can be used for, e.g. research and
CC   nutrition, and may have cytokine and cell proliferation/differentiation,
CC   immune stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC   activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC   receptor/ligand, anti-inflammatory or tumour inhibition activity.
SQ   Sequence   990 BP;    295 A,    197 C,    229 G,    269 T;


alignment_scores:
            Quality:   88.00              Length:    309
              Ratio:  0.620                Gaps:     14
 Percent Similarity:  42.985   Percent Identity: 22.654

alignment_block:
 US-09-174-937-5 x V49593   ..

 Align seq 1/1  to: V49593  from: 1  to: 990

      45 ProGluProPro ..................  ..............As 49
         ||||||||||||                      ||
      88 CCCGAGCCGCCGCCGCTGTCCGGAGCCCCACAGGACGGCATCAGAATTAA 137

      49 nSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaP 66
         |  ||||||||||          ||||||
     138 TGTAACTACACTGAAAGATGATGGG.....       GACATATCTAAAC 175
```

```
  66 rcAsnIleAsnLeuSerLeuAsp GluSer Gly SerIleLeuAspAsp 92
     ||||||  |||||||||      ||||||  ||||||      |||
 176 AGCAGGTTGTTCTTAACATAACCTATGAGAGTGGACAGGTTATTTAAAT 225

  93 Asp          ...          AsnLeuAspSerProAspGlyLeuAs 92
     ||           |||          |||  ||  |||  |||  ||||||
 226 GACTTACCTGTAAATACTGCTGTAACCCCAATAAGGTCTGACAGTTTAAT 275

  93 gIleAsnValAspGluIleAspT rProAspG lAlaAspSerProGluT 113
     ||||||  ||||||||||||||  ||||  |||  |||  |||  |||
 276 AGTGAAGAATGAAAATCTTGAAAATTTCCACGCAAAGAAGTTTTTGAAA 325

 109 yrThrGly..............HisAspProThrAlaAsnLysAspSer 128
     ||||||                |||          ||||||
 326 TTCTCAGTGTAAGGATTTTAGTTCATGAGTGGCCTATGACATCTTTTCC 375

 121 GlyGluGluSerGluSerIleProGluTyrThrAlaGlnGlnArgSer 140
     |||  |||           |||      ||||||  ||  |||  |||
 376 AGTTTGCAACTAATTGTCATTCAACAACACCTACTAGAG..........  414

 137 uAspAsnArgLeuTrpMetThrValValIleGlyGlnGlnSerGlnArgI 154
                              |||  |||  |||  |||  |||
 415 ..........................ATTGATGGAAAACAAGTTCAGAAA 439

 154 leAspMetLysValIleGlu.........ProTyrArgArgValIleSer 167
     ||||||      ||||||          |||  |||  |||  |||
 440 AGGATGTCACTGAAATTGATATTTTAGTTAAGAACCTGGCAGTACTCAGA 489

 169 HisGlyGly                  AspSerGlyTyrTyrGlyAspSer 179
     |||||||||                  |||  |||  |||  |||  |||
 490 CATTCAAACTATACCCTCCCTTTGAAGAAAGTATGTTCTACTCTATTTC 539

 179 yLeuAsnAlaIleIleValPheAlaAlaCysPheIleProAspSerSerA 195
     |  ||||||  | ||||||||||  |  |  |||  |||  |||  |||
 540 TCGAGACAGTGACATTTTATTTACC.........GTTCCTAACCTTCCCA 584

 195 rgAlaAspIyrHisTyrValMetGlAsnLeuPheLeuTyrValIleSer 211
     ||   |||          ||||||||||          |||  |||
 581 AAAAAGAAAGTGTTAGTTCACTGTAAACCACTAGCCAGTATCTTATCAGG 630

 211 ThrLeuGlyLeuMetValAlaGlyAspTyrMetIleValTyrLeuAsnGl 228
     ||||||||      |||  ||      |||  |||  |||  |||  |||
 631 AATGTGGAAACCACTGTAGATGAAGATGTTTTACCTGGCAAGTTA..CC 677

 228 yAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLysCysT 245
     |||||||   |||          |||||||  |||  |||  |||  |||
 678 TGAAACTCCTCTCAGAGCAGAGCCGTTATCTTTATATAAGGTAATTTTC 727

 247 yrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSerPhe...Ile 263
     |||  |||          |||  |||  |||  |||  |||      |||
 728 AGTGGATG   ...GAAAACTTTAGTAAAGATCTGTGTAGGTTCTGAAC 771

 261 IleValHisProSerTrpPhe.............................. 267
     |||  |||  |||  |||  |||
 772 AACGTTTCCCAGTATTCTTTCAGTTTTTGAATCATCATGGTGGTTGGAAT 821

 268 ..........IleArgThrIleLeuAlaValThrArgProLeuPhe 278
             |||  |||  |||  |||  |||  |||  |||  |||
 822 TATATTATCTGTGTTGGTAATATCCATCTTTCATCCTCTTTTTTGTAGTTT 871

 279 IleSerSerLys...PheSerSerLysIleLysTyrValAsnSerLeuSe 294
     ||||||||||||   |||  |||  |||  |||  |||  |||  |||
 872 CTGAATACAAAGGAATTCTTCAGTTGGATAAAGTGGACGTCATATCTGTG 921

 294 rGluLeuSerGlyLeuIleProMet 302
     |||||||||      |||
 922 ACAGCTATCAACTTATATCCAGATG 946
```

```
OM of: US-09-174-937-5 to: GenEmbl:*   out_format : pfs

Date: Feb 9, 2000  4:50 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL frame+_p2n.model -DEV=xlp
 -Q=[illegible]_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.000 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-174-937-5
Query length: 322
Database: GenEmbl:*
Database sequences: 821193
Database length: [illegible]
Search time (sec): 1185.490000


score_list:
Sequence              Strd Orig    ZScore   EScore Len    ! Documentation  ..
gb_pr1:AB002365        + 1601.00 2844.95  1.9e-150  5654  ! AB002365 Human mRNA for KIAA0
gb_pat:AR028453        -  862.50 1530.81  3.0e-77   1380  ! AR028453 Sequence 22 from pat
gb_pr3:HSU15173        -  862.50 1525.33  5.1e-77   2382  ! U15173 Homo sapiens BCL2/aden
gb_ro:AF035207         -  824.50 1457.72  [illegible]  2213  ! AF035207 Mus musculus Nip21 mR
gb_htg3:AC015396       -  479.50  801.09  1.3e-36   78170  ! AC015396 Drosophila melanogas
gb_htg6:AC009556       -  353.00  565.70  1.7e-23   109799  ! AC009556 Homo sapiens chromo
gb_htg6:AC009556       -  308.00  484.76  [illegible]  109798  ! AC009556 Homo sapiens chromo
gb_pr1:HSRHOGAPX       -  [illegible]  [illegible]  [illegible]  1420  ! Z23024 H.sapiens rhoGAP prot
gb_pr1:HSU02570        -  290.00  492.12  [illegible]  3345  ! U02570 Human CDC42 GTPase-act
gb_htg3:AC011488          246.50  373.58  8.7e-13   179301  ! AC011488 Homo sapiens chromo
gb_htg5:AC014962       -  221.50  332.17  1.8e-10   126121  ! AC014962 Drosophila melanoga
gb_pr2:HSU62794        -  157.00  261.86  1.5e-06   1378  ! U62794 Human CDC42 GAP-related
gb_htg1:CEY62F5        -  146.50  188.15  0.0186    310565  ! AL023633 Caenorhabditis eleg
gb_in1:CEW02B12        -  146.50  210.44  0.0011    34305  ! Z66521 Caenorhabditis elegans
gb_in2:AC004293        -  139.00  189.37  0.0181    82095  ! AC004293 Drosophila melanogas
gb_in2:AF091997        -  [illegible]  [illegible]  [illegible]  2234  ! AF091997 Plasmodium falcipar
gb_pl1:SCPPR21         -  122.00  165.88  [illegible]  36017  ! X77688 S.cerevisiae PPR21, FR
gb_pl1:SCYJL201W       -  122.00  194.39  0.0084    2153  ! Z49476 S.cerevisiae chromosome
gb_pl1:YSCPPR21A       -  122.00  194.26  [illegible]  [illegible]  ! [illegible]744 Saccharomyces cerevisi
gb_in2:AF091996        -  [illegible]  [illegible]  [illegible]  1154  ! AF091996 Plasmodium falciparum
gb_ro:MMY17851         -  118.50  186.17  [illegible]  2604  ! Y17851 Mus musculus mRNA for
gb_htg5:AC012719       -  110.00  144.24  5.19      36219  ! AC012719 Drosophila melanogas
gb_pl2:AF082134        -  [illegible]  161.65  [illegible]  4042  ! AF082134 Zea mays copia-like
gb_pl2:AF082130        -  106.00  154.34  1.42      6557  ! AF082130 Zea mays copia-like
gb_pl2:AF082133        -  106.00  154.61  [illegible]  [illegible]  ! AF082133 Zea mays copia-like
gb_vi:HSECOMGEN        -  106.00  122.65  [illegible]  [illegible]  ! M86664 Equine herpesvirus 1
gb_htg2:AC005121       -  105.00  125.85  [illegible]  [illegible]  ! AC005121 Drosophila melanogas
gb_in1:PFMAL3P8        +  105.00  124.11  [illegible]  108908  ! AL034560 Plasmodium falcipar
gb_pl1:POTAGPHI        +  103.50  158.99  [illegible]  2655  ! M69038 Potato alpha-glucan pho
gb_pl1:ATF20M13        -  103.00  122.22  [illegible]  91939  ! AL035540 Arabidopsis thaliana
gb_vi:CV41KBPL         -  102.50  129.50  [illegible]  41997  ! X94355 Cowpox virus 41kbp fra
gb_vi:CVCRI90          -  102.50  127.04  47.17     52283  ! Y11842 Cowpox virus strain GR
gb_pl1:AB010694        -  100.50  130.65  [illegible]  63889  ! AB010694 Arabidopsis thaliana
gb_pl1:AB023035        -  100.50  135.86  15.21     15319  ! AB023035 Arabidopsis thaliana
gb_sts:G37532          -  100.50  172.06  0.1464    428  ! G37532 SHGC-57939 Human Homo sa
gb_ba1:MLCB268         +   99.50  124.64  [illegible]  38859  ! AL022602 Mycobacterium leprae
gb_ba1:MSGB1551CS      -   99.50  125.26  59.21     36548  ! L78813 Mycobacterium leprae c
gb_ba1:MSGB1554CS      -   99.50  125.26  59.21     36548  ! L78814 Mycobacterium leprae c
gb_pl1:AB016892        -   99.50  119.25  [illegible]  66237  ! AB016892 Arabidopsis thaliana
gb_pl2:AF137379        -   99.00  107.12  [illegible]  200799  ! AF137379 Nephroselmis olivac
gb_htg3:AC017451       -   99.00  117.67  [illegible]  70820  ! AC017451 Drosophila melanogas
gb_pl1:SCCHXIV43       -   98.50  121.71  91.43     43481  ! Z46843 S.cerevisiae chromosom
gb_pl1:SCCXIV38K       -   98.50  123.12  77.90     [illegible]  ! [illegible] S.cerevisiae 37.8
gb_pl1:SCYNL126W       -   98.50  [illegible]  [illegible]  [illegible]  ! Z71402 S.cerevisiae chrom
gb_pl1:SCYNL127W       -   98.50  [illegible]  [illegible]  [illegible]  ! Z71403 S.cerevisiae chrom

seq_name: gb_pr1:AB002365

seq_documentation_block:
LOCUS       AB002365     5654 bp    mRNA            PRI       [illegible]-FEB-1999
DEFINITION  Human mRNA for KIAA0367 gene, partial cds.
ACCESSION   AB002365
VERSION     AB002365.1  GI:2224674
KEYWORDS    KIAA0367.
SOURCE      Homo sapiens male brain [illegible]
            plus clone:HH0141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (2?-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure I; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:[illegible]@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3[illegible],
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349384
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescript [illegible]"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA0367"
     CDS             <1..2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20822.1"
                     /db_xref="GI:2224675"
                     /translation="[illegible]HSFSDASG
                     LNISIGIIDD[illegible]NQNIDEDLL
                     EFEHFLY[illegible]HRAGSPEDESLGC
                     RAAEIVLSALFDR[illegible]JASSNILSFSN
                     VDWEVIDINSLL[illegible]LASFPDTC
                     EDRHALHMDYILVN[illegible]ARDSSPGWS
                     [illegible]
                     [illegible]AJEIAVIH
                     DSEWEML[illegible]PFEEGVLS
                     PSAADMRPEPPNS[illegible]NLISPDEIDINV
                     DELDIHDEADS[illegible]VIGEQEQR
                     IDMFVIEP[illegible]LFLYVISTLE
                     LMVAELYNIVI[illegible]"
BASE COUNT     1664 a   1188 c   [illegible] g   [illegible] t
ORIGIN

alignment_scores:
           Quality: 1601.00            Length:    [illegible]
             Ratio: 5.035                Gaps:    [illegible]
 Percent Similarity: 90.341    Percent Identity: [illegible]

alignment_block:
 US-09-174-937-5 x AB002365  ..
```

```
Align seq 1/1  to: AB002365  from: 1  to: 5654

      1 MetGluGluGluThrGluPheLeuGluLeuGlyThrArgIleSerArgPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1414 ATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGAACCAGGATATCAAGACC 1463

     17 oAsnGlyLeuLeuSerGluAspValGlyMetAspIleProPheGluGluG 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1464 AAATGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGG 1513

     34 lyValLeuSerProSerAlaAlaAspMetArgProGluProProAsnSer 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1514 GCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCT 1563

     51 LeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaProAs 67
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   1564 CTGGATCTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAA 1613

     67 nIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspA 84
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   1614 TATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATA 1663

     84 snLeuAspSerProAspGluIleAspIleAsnValAspGluLeuAspThr 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1664 ACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACC 1713

    101 ProAspGluAlaAspSerPheGluTyrThrGlyHisAspProThrAlaAs 117
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1714 CCCGATGAAGCAGATTCTTTTGAGTACACTGGCCATGATCCCACAGCCAA 1763

    117 nLysAspSerGlyGlnGluSerGluSerIleProGluTyrThrAlaGluG 134
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   1764 CAAAGATTCTGGCCAAGAGTCAGAGTCTATTCCAGAATATACCGCCGAAG 1813

    134 luGluArgGluAspAsnArgLeuTrpMetThrValValIleGlyGluGln 150
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   1814 AGGAACGGGAGGACAACCGGCTTTGGATGACAGTGGTCATTGGAGAACAA 1863

    151 GluGlnArgIleAspMetLysValIleGluProTyrArgArgValIleSe 167
        |||||||||||||||||||||||||||| |||||||| ||||||||||||
   1864 GAGCAGCGCATTGACATGAAGGTCATCGAGCCCTACAGGAGAGTCATTTC 1913

    167 rHisGlyGlyAspSerGlyTyrTyrGlyAspGlyLeuAsnAluIleIleV 184
          |||           |||||||||||||||||||||||||||||||||
   1914 TCACGGA         GGATACTATGGGGACGGTCTAAATGCCATCATTG 1954

    184 alPheAlaAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyr 200
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   1955 TGTTTGCCGCCTGTTTTCTGCCAGACAGCAGTCGGGCGGATTACCACTAT 2004

    201 ValMetGluAsnLeuPheLeuTyrValIleSerThrLeuGluLeuMetVa 217
        |||||||||||||||||||||||||||| |||||||||||||||||||||
   2005 GTCATGGAAAATCTTTTCCTATATGTAATAAGTACTTTAGAGTTGATGGT 2054

    217 lAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThr.ProArgArg 233
        ||| ||| |||||||||| |||||||||| || |||||| | |||||||
   2055 AGCTGAAGACTATATGATTGTGTACTTGAATGGTGCAACCCCCAAGAAGG 2104

    234 ArgMetProGlyLeuGlyTrpMetLysLysCysTyrGlnMetIleAspAr 250
        |||||||||||||||||||||||||||||| |||||||| |||||||||
   2105 AGGATGCCAGGGCTAGGCTGGATGAAGAAAATGCTACCAGATGATTGACAG 2154

    250 gArgLeuArgLysAsnLeuLysSerPheIleValHisProSerTrpP 267
        |||||||||||||||||||||||||||||| |||||||||||||||||
   2155 AGGGTTGAGGAAGAATTTGAAATCATTCATCATTGTTCATCCATCTTGGT 2204

    267 heIleArgThrIleLeuAlaValThrArgProPheIleSerSerLysPhe 283
        |||||||||||||||||||||||||||||| |||||||||||||||||||
   2205 TCATCAGAACAAATCTTGCTGTGACAGCACTTTATAAGTTCAAAATTC 2254

    284 SerSerLysIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIl 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||

      2255 AGCAGTAAAATTAAATATGTCAATAGCTTATCAGAACTCAGTGGGCTGAT 2304

       300 eProMetAspCysIleHisIleProGluSerIleIle..... 312
           |||||||||||||||||||||||||||| | |||||||
      2305 CCCAATGGATTGCATCCACATTCCAGAGAGCATCATCAAACTGGATGAAG 2354

       312 ..... 312

      2355 AACTGAGGGAAGCATCAGAGGCAGCTAAAACTAGCTGCCTTTACAATGAT 2404

       313 .....AsrIleAspLeuLysLeuLysGluLy 321
                ||||||||||||||||||||||||||||
      2405 CCAGAAATGTCTTCTATGGAGAAGGATATTGACTTGAAGCTGAAAGAAAA 2454

       321 sPro 322
           ||||
      2455 GCCT 2458

seq_name: gb_pat:AR028453

seq_documentation_block:
LOCUS       AR028453    1386 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 22 from patent US 5858678.
ACCESSION   AR028453
VERSION     AR028453.1  GI:5940426
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G.
  TITLE     Apoptosis-regulating proteins
  JOURNAL   Patent: US 5858678-A 22 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    346 g    373 t
ORIGIN


alignment_scores:
            Quality:  862.50            Length:    301
              Ratio:  3.549               Gaps:      6
 Percent Similarity:  80.731   Percent Identity: 58.472

alignment_block:
 US-09-174-937-5 x AR028453

 Align seq 1/1  to: AR028453  from: 1  to: 1386

      23 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSe 39
         ||||||      :::   :::|||   ::::::::::::   :::
     250 GAAGATTTTCCGATACCTTTACCAGAAGATGATAGTATTGAAGCAGATAT 299

      39 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 55
           |||          ||||||      |||:::|||||||:::::::'
     300 ACTAGCTATAACTGGACCAGAGGACCAGCCTGGCTCACTAGAAGTTAATG 349

      55 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 71
         :::    ||':::   ||||||   |'||||:::|.|:::   :::
     350 GAAATAAAGTGAGAAAG...AAACTAATGGCTCCAGACATTAGCCTGACA 399

      72 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 88
         ||||||   |||   ||||||  : | | | |'|||   :|:.:.:::
     397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 443

      88 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGln....A 103
          ||||||||||:::      |||   '| ||||| ||':::      :
     444 TGGGGAGATTGACTTA    .GATGGCTTAGACACAGCGTCAGAGAATA 487

     104 laAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspSer 120
         .....:|||||||:::          |:|     :::      :::
     488 GTAATGAGTTTGAGTGGGAAGATGATCTTCCAAAACCCAAGACTACTGAA 537
```

```
     121 GlyGlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgGl 137
     538 GTAATTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGA 587

     137 uAspAsnArgLeuTrpMetThrValValIleGlyGluGlnGluGlnArgI 154
     588 AGATGGACGACGCTGGCGTATGTTCAGGATTGCAGAACAGGACCACAGGG 637

     154 leAspMetLysValIleGluProTyrArgArgValIleSerHisGlyGly 170
     638 TTGATATGAAGGCAATTGAACCCTATAAAAAACTTATCAGCCATGGG... 684

     171 AspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAl 187
     685 .     GGATATTATGGGGATGGATTAAATGCCATTGTTGTATTTGCTGT 728

     187 aCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluA 204
     729 CTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAGATACCTGATGGACA 778

     204 snLeuPheLeuTyrValIleSerThrLeuGluLeuMetValAlaGluAsp 220
     779 ATCTTTTTAAATATGTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAAC 928

     221 TyrMetIleValTyrLeuAsnGlyAlaThrIleArgArgArgMetProGl 237
     829 TACATGATAGTTTATTTAAAATGGTGCAACAACTCGAAGAAAAATGCCCAG 878

     237 yLeuGlyTrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgL 254
     879 TCTGGGATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGA 928

     254 ysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThr 270
     929 AAAATCTAAAATCCCTAATCATTGTACATCCTCTTGGTTTATCAGAACA 978

     271 IleLeuAlaValThrArgProPheIleSerSerLysPheSerSerLysIl 287
     979 CTTCTGGCTGTTACAAGACCATTTATTAGCTCAAATTCAGCCAAAAAAT 1028

     287 eLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspC 304
    1029 TAGATACGTGTTTAAATTTGGCAGAACTAGCAGAACTTGTCCCCATGGAAT 1078

     304 ysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeuLysGlu 320
    1079 ACGTTGGCATACCAGAATGCATAAAACAAGTTCATCAAGAACTTAATGGA 1128

     321 Lys 321
    1129 AAA 1131

seq_name: gb_pr3:HSU15173

seq_documentation_block:
LOCUS       HSU15173     2382 bp    mRNA            PRI       02-FEB-1998
DEFINITION  Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2)
            mRNA, complete cds.
ACCESSION   U15173
VERSION     U15173.1  GI:559843
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 212 to 1156)
  AUTHORS   Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K.,
            Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.
  TITLE     Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
            of cellular proteins
  JOURNAL   Cell 79, 341-351 (1994)
  MEDLINE   95042730

REFERENCE   2  (bases 1 to 2382)
  AUTHORS   Boyd,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1994) James M. Boyd, Institute for Molecular
            Virology, St. Louis University Health Sciences Center, 3681 Park
            Avenue, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B cell, S.Elledge; B-cell, J.Ambrus; fetal
                     brain, M.Green"
                     /cell_type="B cells; brain"
                     /dev_stage="fetus (brain)"
     gene            1. .2382
                     /gene="BNIP2"
     CDS             212. .1156
                     /gene="BNIP2"
                     /codon_start=1
                     /product="BCL2/adenovirus E1B 19kD-interacting protein 2"
                     /protein_id="AAC50121.1"
                     /db_xref="GI:559844"
                     /translation="MESVELKEEWLDELFPIPLPEEISEAFIIAITGFEDQPGSLEV
                     NGNKVRKKLMAFLSLLLPSLSSLSLLLLSSSSSSSSSSSSSSSSSSSSSSSSSSNS
                     KPFTTTVIPKSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     GDGLNAIVVPAVTPMPSSPQPNVPVLMENLPKVVIGLLLVAFNYMIVLNGATTRR
                     KMPSLGWIPKCYLLSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
                     NLAELAELVPMEYVSIPEDIK,VDLELNSKLEFKNPL"
BASE COUNT      734 a    393 c    511 g    744 t
ORIGIN


alignment_scores:
           Quality:  862.50              Length:    9 1
             Ratio:  3.549                 Gaps:      6
 Percent Similarity:  80.731   Percent Identity:  88.472

alignment_block:
 US-09-174-937-5 x HSU15173   ..

 Align seq 1/1  to: HSU15173  from: 1  to: 2382

      23 GluAspValGlyMetAspIlePr.PheGl.GlGlyVal.uSerProSe 64
     248 GAAGATTTTCCGATACCTTTACCAGAAGATGATAGTATTGAAGTAGATAT 297

      39 rAlaAlaAspMetArgProGl....P.uP.PLA.uSerLeuAspL.AspA 88
     298 ACTAGCTATAACTGGACCAGAGGACCAGTCGTGGTCACTAGAAGTAATG 347

      55 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 104
     348 GAAATAAAGIGAGAAAG...AAACTAATGGTTCTAGACATTAGCTTGACA 394

      72 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 88
     395 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 441

      88 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspG.....A 1.4
     442 TGGGGAGATTGACTTA.....GATGGTTAGATACACTGTTAGATAAATA 485

     104 laAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspSer 121
     486 GTAATGAGTTTGAGTGGGAAGATGATCTTCCTAAAACCCAAGACTACTTAA 535

     121 GlyGlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgGl 137
     536 GTAATTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGA 585

     137 uAspAsnArgLeuTrpMetThrValValIleGlyGluGlnGluGlnArgI 154
     586 AGATGGACGACGCTGGCGTATGTTCAGGATTGCAGAACAGGACCACAGGG 635
```

```
    154 leAspMetLysValIleGluProTyrArgArgValIleSerHisGlyGly 170
        ::''''''''::::|||||''''|||:::::::  |''''''''|||||
    636 ITGATATGAAGGCAATTGAACCCTATAAAAAAGTTATCAGCCATGGG    682

    171 AspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAl 187
          '|||||||||||''''''''''''''''''.||  ||:::||||||||::
    683 ......GGATATTATGGGGATGGATTAAATGCCATTGTTGTATTTGCTGT 726

    187 aCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluA 204
        :|||||||:: :::  :|||||        :::  : |||   |||  :|
    727 CTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAGATACCTGATGGACA 776

    204 aIleLeuPheLeuTyrValIleSerThrLeuSerLeuMetValAlaGluAsp 220
        ||||||||   ||||||||||:::||||||||||:|  :|||||||||:.
    777 ATCTTTTTAAATATCTTATTGGCACTTTGGAGCCATTAGTAGCAGAAAAC 826

    221 TyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetProGl 237
        ||||||:'''''''''''''''''''''''''    ||||:::|||||||::
    827 TACATGATAGTTTATTTAAATGGTGCAACAACCTGAAGAAAAATGCCCAG 876

    237 yLeuGlyTrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgL 254
        :''''''':|:::::::|||||||||||||:::|  ||||||||||||||||
    877 TCTGGGATGGCTCAGGAAATGTTATCAGCAAA'GATAGAAGGTTACGGA 926

    254 ysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThr 270
        ||||||||||||||:::||||||||||| ||' ||||||||||||||||
    927 AAAATCTAAAATGCCTAATCATTGTACATCCT'CTTGGTTTATCAGAACA 976

    271 IleLeuAlaValThrArgProPheIleSerSerLysPheSerSerLysIl 287
        :::|||||||||||||||||||||||||||'''':|||||||||:::|||||
    977 CTTCTGGCTGTTACAAGACCATTTATTAGCTCGAAATTCAGCCAAAAAAT 1026

    287 eLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspC 304
        |:::|||||   :::|||:::||||||:::   |||...|||:|||...
   1027 TAGATACGTGTTTAATTTGGCAGAACTAGTAGAAGTTGTCGCCATGGAAT 1076

    304 ysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeuLysGlu 320
          :::   '|'|'|'||   '||   :::::  :   :::|||:::
   1077 ACGTTGGCATACCAGAATGCATAAAACAAGTTGATCAAGAACTTAATGGA 1126

    321 Lys 321
        ||
   1127 AAA 1129
```

```
seq_name: gb_ro:AF035207

seq_documentation_block:
LOCUS       AF035207     2213 bp    mRNA            PRI       25-FEB-1999
DEFINITION  Mus musculus Nip2l mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip2l gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, CuraGen Corporation,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1. .2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit54 and D9Mit1"
     gene            1. .2213
                     /gene="Nip2l"
     CDS             134. .1114
                     /gene="Nip2l"
                     /note="associates with motifs of both the adenovirus E1B
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP2l"
                     /protein_id="AAC04329.1"
                     /db_xref="GI:2911348"
                     /translation="MPGVELKPEWQDPDPFIPLPEDDGIEADILDGTDPDRQPGSLEV
                     NGNKVPKKLMAPDISLTLDPGPGSLWSDELDRAGEVDLPQLDTPSPNSTPFEWEDDLP
                     KPKTTEVIRKGSIISYIATEPPQDQPPWPMFPIEQDPRVIMKAIPPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQPNYPYLMCNLPPYVIGTLELIGMENYMIIVLNGATIPR
                     KMPSLGWLRPCYQQIDPRLRKNLKSLTTVHPSWFIRTIIAVTRPFLSSPFSQKIRYVF
                     NLAELAELVPMPYYVIPPCIPQYEFPPPKPKPVPQFLNGPPPPSEQ"
BASE COUNT      637 a    400 c    571 g    605 t
ORIGIN


alignment_scores:
           Quality:  824.50            Length:    316
             Ratio:  3.393               Gaps:      7
 Percent Similarity:  76.899   Percent Identity: 53.165

alignment_block.
 US-09-174-937-5 x AF035207   ..

 Align seq 1/1  to: AF035207  from: 1  to: 2213

     23 GluAspValGlyMetAspIleProPheGluGluGlyValLeuSerProSe 69
        ||||||      :::   :::|||   ::::::::::::   :::   ::
    170 GAAGATTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATAT 219

     39 rAlaAlaAspMetArgProGlu...ProProAsnSerLeuAspLeuAsnA 85
        :               ||'''      |||::|||||:::::::||.
    220 ACTAGATGGAACTGATCCAGACAGACAGCCTGGCTCCTTAGAAGTTAATG 269

     55 spThrHisProArgArgIleLysLeuThrAlaProAsnIleAsnLeuSer 71
        :::      |||:::  ||||'|  ||||||:::|||:::|||:::
    270 GGAACAAAGTAAGGAAG...AAACTGATGGCCCCAGACATCAGCCTGACC 316

     72 LeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAspSerPr 88
        ||||||   :::|||||||||||:::  '|''|| ||   |'|:|:::
    317 CTGGATCCTGGTGAAGGTTCTCTGTGGTCCGATGAT...TTGGATGAAGT 363

     88 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 104
        |||:::|||:::      :::  |:|||.|||||||:::'''  :
    364 TGGAGAGGTTGACCTG      GAAGGCTTAGACAGCCCATCAGAGAAGA 413

    104 laAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspSer 120
        ::|||:::|||||''':           |||      :::      :::
    408 GTGATGAGTTTGAGTGGGAAGATGATCTTCCCAAACCCAAAACTACTGAA 457

    121 GlyGlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgIl 137
          ..  ||||||  ||||||:||||   ||:| |:::
    458 GTCATTAGGAAAGGCTCGATTACTGAGTATACAGCCGACAGAAGAAAAGG 507

    137 uAspAsnArgLeuTrpMetThrValIleGlyGluGluGluGlnArgI 154
         |||...|||   |||          :: ||||||||||||:::: ||
    508 CGATGGACGCCGCTGGCGCATGTTCAGGATTGGAGAGCAGGACCACAGG 557

    154 leAspMetLysValIleGluProTyrArgArgValIleSerHisGlyGly 170
        :'|||||||:   ||||||||||'':::::'||||'|'|||'''  :
    558 IGGACATGAAGGCGATCGAGCCCTACAAAAAAGTTATTAGCCATGGA... 604

    171 AspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAl 187
        |||||||||||||||||||||||||||||||||||:::||'''  ::
    605     GGATATTATGGGGACGGCTTAAATGCCATTGTTGTATTTGCTGT 648

    187 aCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluA 204
        :|||||:::||:::|||:::::   :::||||:::|:::|||  :::.
    649 CTGTTTCATGCCTGAGAGTGGTCAGCCTAACTATAGATATTTGATGGACA 698
```

```
seq_name: gb_htg5:AC015396

seq_documentation_block:
LOCUS       AC015396    78170 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015396
VERSION     AC015396.1  GI:6435939
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78170)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209602 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1   78170
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21762 a  17118 c  17312 g  21978 t
ORIGIN


alignment_scores:
            Quality:  479.50            Length:    321
              Ratio:  2.510               Gaps:     12
 Percent Similarity:  59.502   Percent Identity:  36.449

alignment_block:
 US-09-174-937-5 x AC015396/rev   ..
```



```
Align seq 1/1  to reverse of: AC015396  from: 1  to: 78170
```

```
    8192 CAAAGTTTTGGGCAAAGCTGGTCTACGTAAAGTCCCTGGAGGAACTGGGC 8143

     298 GlyLeuIleProMetAspCysIleHisIleProGluSerIleIleAsnIl 314
         ...  ......          |||||  ||:::::   ...
    8142 ATGCAGGTGGTCGTGGAGAAGGCAGGCATCCCCGAGAAGGTCAAGCAATA 8093

     314 eAspLeuLys 317
         |||  |||
    8092 CGACGCCAAG 8083

seq_name: gb_htg6:AC009556

seq_documentation_block:
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cook,P., DeArellano,K., Dupaple,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Levine,J., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Mella,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Savery,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yo,W.T. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.genome.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: LL67L
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *              gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *              gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *              gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *              gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *              gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *              gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *              gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *              gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *              gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *              gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *              gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *              gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *              gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *              gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *              gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *              gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *              gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *              gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *              gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *              gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *              gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *              gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *              gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *              gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *              gap of unknown length
            *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


alignment_scores:
            Quality:  353.00              Length:     67
              Ratio:  5.269                 Gaps:      0
 Percent Similarity:  100.000   Percent Identity:  100.000
```

```
alignment_block:
 US-09-174-937-5 x AC009556   ..

 Align seq 1/1  to: AC009556  from: 1  to: 169798

      46 GluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLy 62
         |||||||||||||||||||||||||||||| |||||||||||||||||||
  119825 GAACCTCCTAATTCTCTGGATCTTAAIGACACTCATCCTCGGAGAATCAA 119874

      62 sLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerI 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  119875 GCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 119924

      79 leLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsnVal 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  119925 TTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTG 119974

      96 AspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGlyHi 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
  119975 GATGAACTTGATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGGCCA 120024

     112 s 112
         |
  120025 T 120025

seq_name: gb_htg6:AC009556

seq_documentation_block:
LOCUS       AC009556   169798 bp    DNA              HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P , DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D , Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liew,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M , O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,P , Roberts,D., Roy,A , Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A ,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H , Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100%  1 reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 120385 bases at least 14.
                Consensus quality: 156543 bases at least 18.
                Consensus quality: 165514 bases at least 20.
                Insert size: 165000; agarose-ip
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in 620 bases; agarose-ip
                Quality coverage: 4.1 in 641 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1016: contig of 1016 bp in length
            *               gap of unknown length
            *      1017     2036: contig of 1020 bp in length
            *               gap of unknown length
            *      2037     3371: contig of 1335 bp in length
            *               gap of unknown length
            *      3372     4972: contig of 1601 bp in length
            *               gap of unknown length
            *      4973     5981: contig of 1009 bp in length
            *               gap of unknown length
            *      5982     8008: contig of 2027 bp in length
            *               gap of unknown length
            *      8009     9286: contig of 1278 bp in length
            *               gap of unknown length
            *      9287    11243: contig of 1957 bp in length
            *               gap of unknown length
            *     11244    14312: contig of 3069 bp in length
            *               gap of unknown length
            *     14313    18423: contig of 4111 bp in length
            *               gap of unknown length
            *     18424    23065: contig of 4642 bp in length
            *               gap of unknown length
            *     23066    30558: contig of 7493 bp in length
            *               gap of unknown length
            *     30559    37537: contig of 6979 bp in length
            *               gap of unknown length
            *     37538    44831: contig of 7294 bp in length
            *               gap of unknown length
            *     44832    50944: contig of 6113 bp in length
            *               gap of unknown length
            *     50945    58263: contig of 7319 bp in length
            *               gap of unknown length
            *     58264    65216: contig of 6953 bp in length
            *               gap of unknown length
            *     65217    72310: contig of 7094 bp in length
            *               gap of unknown length
            *     72311    78799: contig of 6489 bp in length
            *               gap of unknown length
            *     78800    87607: contig of 8808 bp in length
            *               gap of unknown length
            *     87608    98337: contig of 10730 bp in length
            *               gap of unknown length
            *     98338   112198: contig of 13861 bp in length
            *               gap of unknown length
            *    112199   126330: contig of 14132 bp in length
            *               gap of unknown length
            *    126331   140617: contig of 14287 bp in length
            *               gap of unknown length
            *    140618   153843: contig of 13226 bp in length
            *               gap of unknown length
            *    153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


alignment_scores:
            Quality:  308.00             Length:    291
              Ratio:  3.899                Gaps:      1
 Percent Similarity:  27.148   Percent Identity:  27.148

alignment_block:
 US-09-174-937-5 x AC009556/rev  ..

 Align seq 1/1  to reverse of: AC009556  from: 1  to: 169798

     173 GlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 189
         | | ||||||||||||||||||||||||| || |||||||||| |||||||
   13902 GCATACTATGGGCACGGTCTAAATGCCATCATTGTTTGCCGCCTGTTT 13853

     189 eLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuP 206
         |||||||||||||||||||||||||||||||| ||||||||||||||||
   13852 TCTGCCAGACAGCAGTCGGGCGGATTACCACTATGTCATGGAAAATCTTT 13803

     206 heLeu                          ... .. ..... ..... 207
         |||||
   13802 TCCTCTGACTCATACTTATGACAAAGTGCTTTGAGTTAATGTCTGTGGCA 13753

     207                                 ...      .  ...  . 207

   13752 TGTAATGGCAGATAGAGCACCTTTAATTAAAGAAAATAGATGATGTAGAA 13703

     207                                          ..... ... 207

   13702 TCTTCTAAGCCTTCAGTGAAAAGTCAGTAAGG   TCTGTCAAGGAATGTT 13653

     207 ..........................        .  .  ... ... 207

   13652 GTTCAATGGGGTAAAATCTTTAAAGGAGCTCTT TGTGGGTAGACTAGAC 13603

     207 ....................                 .  .  ..... 207

   13602 TGAAACCAAGATGAACAGAATTTTGCTTGCAT  TAAAATTCTGTAACAT 13553

     207                              .     ..... ..... 207

   13552 TTCCTTTGGGCAAAGTAITATTTTTGTTCCTTCATAACATGTTCTTAGTC 13503

     207 ....................                 .  .  ..... 207

   13502 AGCTAAGACTGACTATAACCTGACTTCATTAATATCTGCCCTTATTTAAC 13453

     207                                                  207

   13452 CTGAGTTACCCTTAGTATTTGATAACATTGTTTACAGTTTTGAAAGTCAG 13403

     207                                      ... ..... ... 207

   13402 GATTCTTTTTAACCCATACATTTTTTCTAGTCTTTGTCCATTTAAAAGATA 13353

     207 ..................................................... 207

   13352 GATAGCCATACCCGATAAATGAGTAATACTCTCATAGGAATGCTCTTAGGT 13303

     207 ..........................          .  .  ... 207

   13302 AATTGTATTCTTCCCAAGGTAAAGAGGTAATATTCACAATCATTCTTAAA 13253

     207 ..................................................... 207

     .
```

```
   13252 TTTCATACACCCTGTACTTTTGATCTTATGTAATTATGTCTGTGAATACCA 13203

     208 ..........................  ..............TyrValIle 21
                                                    | | ||||

   13202 TGTGGAACTGAAAACATTCCTTTTTTTTTTTGATCAGATACTTAATA 13157

     211 SerThrLeuGluLeuMetValAlaGluAspTyrMetIleValTyrLeuAs 227
         |||||||||||||||||||||||||| |||||||||||||| || | |
   13156 AGTACTTTAGAGTTGATGGTAGCTGAAGACTATATGATTGTGTACTTGAA 13107

     227 nGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLysC 244
         |||||||||||||||||||||||||| |||||||||||||| |||||||
   13106 TGGTGCAACCCCAAGAAGGAGGATGCCAGGGTAGGCTGGATGAAGAAAT 13057

     244 ysTyrGlnMetIleAspArgArg 251
         |||||||||||||||||||||||
   13056 GCTACCAGATGATTGACAGACGG 13033

seq_name: gb_pri:HSRHOGAPX

seq_documentation_block:
LOCUS       HSRHOGAPX    1430 bp    mRNA            PRI       25-AUG-1997
DEFINITION  H.sapiens rhoGAP protein.
ACCESSION   Z23024
VERSION     Z23024.1  GI:312211
KEYWORDS    rhoGAP protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2): 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   Hall,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1993) Alan Hall Dr.  Cell and Molecular Biology,
            Institute of Cancer Research, Chester Beatty Laboratories, 239
            Fulham Road, London, SW3 6JB, England
FEATURES             Location/Qualifiers
     source          1..1430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="Male"
                     /tissue_type="fibrosarcoma"
                     /cell_line="HT1080"
     5'UTR           1..98
     CDS             99..1418
                     /codon_start=1
                     /product="rhoGAP protein"
                     /protein_id="CAA80560.1"
                     /db_xref="GI:312212"
                     /db_xref="SWISS-PROT:Q07960"
                     /translation="MLPLSELGDDLTLDDTSEALNQVKLAS[illegible]KNWISEMPDFPKS
                     DDSKSSSPELVTHLKWDDPYYEIAPHQIVEVAGLDRYSPKEIVESAAEMIPSHQLDHS
                     KLLGYLKHTLDQYVESDYTLLYLHHGLTSDNKPPSLSWLPDAMSPDDPKYKKNIKALYI
                     VHPTMPIKTLLILFKPLISPFPGQKIFYVNYLSELSPHVKFEL[illegible]TPE_VLKYDDFLK
                     STQKSPATAPRFMSFAFLPNQLPGVSLLHLQERNHPDEFTIYNLRDVNYLQAHALT
                     TEGIFRRSANTQVVRFVQQKYNMGIPVDFDQYNFLHLPAVTLKFPLPELPELLTFDL
                     YPHVVGFLNIDESQFVPATLQVLQTLPEHTQVLRFLTAFLVLIAAPY_NKMTNTNL
                     AVVFGPNLLWAKDAAITLKAINPINTFTKFLLDHQGELPPSLTTSSL"
BASE COUNT      332 a    453 c    395 g    280 t
ORIGIN


alignment_scores:
            Quality:  290.00             Length:    267
              Ratio:  1.847                Gaps:      7
```

```
 Percent Similarity:  58.801   Percent Identity: 27.341

alignment_block:
 US-09-174-937-5 x HSPHOGAPX   ..

 Align seq 1/1  to: HSPHOGAPX  from: 1  to: 1430

      60 ArgIleLysLeuThrAlaProAsnIleAsnLeuSer...LeuAspGlnSe 75
         |||   ::::|||.:: :::   .:::::     ::: |||:::.:::
      69 AGGGCAAGGCTGTCTGGGACAGGCTTGGCCATGGATCCGCTCTCAGAGCT 118

      75 rGluGlySerIleLeuSerAspAspAsnLeuAspSerProAspGluIleA 92
         :::   ::::::      ||||||:::  :::::     : ::::
     119 GCAGGATGATCTGACCTTGGATGACACCAGCGAGGCTCTGAACCAGCTGA 168

      92 spIle...AsnValAspGluLeuAspThrProAspGluAlaAspSerPhe 107
           :::   :.....||||||   .::   |||
     169 AGCTGGCCTCCATCGATGAGAAGAACTGGCCC..................... 200

     108 GluTyrThrGlyHisAspProThrAlaAsnLysAspSerGlyGlnGluSe 124
                                                          ||
     201 ...  .  .  .  .  .  .  .  .  .  .  .  .  .  .....TC 202

     124 rGluSerIleProGluTyrThrAlaGluGluGluArgGluAspAsnArgL 141
         |.....  ||:::.:.    :::    .  .  .  :::
     203 GGATGAAATGCCTGACTTCCCCAAGTCAGATGACTCCAAAAGCAGC.... 248

     141 euTrpMetThrValValIleGlyGluGlnGluGlnArgIleAspMetLys 157
                                 :::   |||             ::::||
     249 .  .  .  .  .  .  .  .  .  .TCCCCGAACTTGTCACACACCTGAAG 275

     158 ValIleGluProTyrArgArgValIleSerHis..  .  .  .  .  .Gl 169
             :::::::::   .:::    :::                      ::
     276 TGGGATGACCCATACTATGACATCGCCCCGCAGTCAGATCGTGGAGGTGGC 325

     169 yGlyAspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheA 186
         :::::::::   |||               :::   ||||||||||||:
     326 AGGAGATGACAAGTATGGG..      .....CCGAAGATCATTGTGTTTA 362

     186 laAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMet 202
         ::::::::   ...|||   |||   ...   |||  ...
     364 GTGCCTGTCGAATGCCCCCAGCCACCAGCTCGACCAC............. 401

     203 GluAsnLeuPheLeuTyrValIleSerThrLeuGluLeuMetValAlaGl 219
         ::::::::::::   :|||:::      |||||:::      |||   ::
     402 AGCAAGCTCCTGGGGTACCTGAAGCACACCCTCGACCAGTACGTGGAGAG 451

     219 uAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetP 236
         :|||||   ::  . |||||..                        ...  |
     452 TGACTACACAGTTCTGTATCTGCACCACGGCCTGACCAGCGACAACAAGC 501

     236 roGlyLeuGlyTrpMetLysLysCysTyrGlnMetIleAspArgArgLeu 252
         ||:::|||:::|||. .  .  .  .|||      .:|||||'
     502 CCTCCCTCAGCTGGCTCCGTGATGCCTACCGGCAGTTTGACCGCAAGTAC 551

     253 ArgLysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleAr 269
         ..:|||||:..|||.....  ||||||'|'|||::   |||||||::
     552 AAGAAGAACATCAAGGCCTTGTACATCGTGCATCCAACCATGTTCATCAA 601

     269 gThrIleLeuAlaValThrArgProPheIleSerSerLysPheSerSerL 286
         :|||...|||   .::   ..:|||.:|'|||'   |||||||:::::|
     602 AACTCTGCTCATCCTCTTCAAGCCCCTCATCAGCTTCAAGTTCGGGCAGA 651

     286 ysIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMet 302
         |||||  ||||||||||  ||||||||||||||||     :::   ...
     652 AGATCTTCTATGTGAATTACCTGAGCGAGCTGAGCGAGCACGTGAAGCTG 701

     303 AspCysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeuLy 319
         ::   ::   |||||:.:::::::::::::::::  |||      |||||
     702 GAGCAGCTGGGGATCCCTCGCCAAGTGCTCAAATATGACGACTTCCTGAA 751

     319 s 319
         |
```

```
     752 A 752

seq_name: gb_pri:HSU02570

seq_documentation_block:
LOCUS       HSU02570     3345 bp    mRNA            PRI       27-JAN-1994
DEFINITION  Human CDC42 GTPase-activating protein mRNA, partial cds.
ACCESSION   U02570
VERSION     U02570.1  GI:450780
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3345)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 268, 26159-26165 (1993)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 3345)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P., Self,A.J., Brill,S., van Erp,H.E.
            and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase activating protein for
            Rho-Related Small GTPases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3345)
  AUTHORS   Barfod,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1993) Elisabeth T. Barfod, Genetech, Inc., 460
            Point San Bruno Blvd., South San Francisco, CA 94080, USA
COMMENT     On Jan 27, 1994 this sequence replaced gb|L25080|HSP50GAP.
FEATURES             Location/Qualifiers
     source          1. .3345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CMK"
     CDS             <1. .1391
                     /codon_start=3
                     /function="GTPase activating protein for the rho family of
                     small GTP-binding proteins that is most effective against
                     CDC42"
                     /product="CDC42 GTPase-activating protein"
                     /protein_id="AAA16142.1"
                     /db_xref="GI:450781"
                     /translation="QSSAPAFPLLPPCPAPLSITGLAMDPLSPLCCLTLDDTSEALN
                     CLPLAITMDPL[illegible]ASPLSPeeeLL[illegible]AHeQIVEVAGDD
                     KYGPKILVPSAQRMFPSWLLGMGKILSVLRNTIEQYVESL[illegible]
                     SWLRIAYREFDPVPRRNIALLLVFTNIAAL[illegible]YVNYLSEL
                     SEHVKIPQLPIPLVLKPPDPLKSIHSPAAMRPMFPFTLNGLQHLQEKN
                     PEQLPIPIVLFELVAELAHAHIISGIPPSANIAVVPLVLKLNMSDVDFDQYNEL
                     HLPAVILKTFLRELPDPLLTYDLYPHVVSFLNIDERKRVPATLVQLVLPEENYQVLR
                     FLTAFLVQIGPHSDQNRMTNTMLAVVFGPNLLWARQAAILAAGLLSVIKFLLDHQ
                     GELFPSPDPSGL"
     misc_feature    753. .785
                     /note="SH3-binding domain; capable of binding the SH3
                     domains of src and p85 alpha; aa residues 251-261"
     misc_feature    849. .926
                     /note="active site box 1; aa residues 283-308"
     misc_feature    999. .1091
                     /note="active site box 2; aa residues 333-363"
     misc_feature    1149. .1277
                     /note="active site box 3; aa residues 384-425"
     polyA_signal    3328. .3333
     polyA_site      3345
                     /note="38 A residues"
BASE COUNT      655 a   1061 c    855 g    774 t
ORIGIN


alignment_scores:
          Quality:  290.00            Length:    2[illegible]
```

```
            Ratio:  1.847                Gaps:      7
 Percent Similarity:  59.901   Percent Identity: 27.341

alignment_block:
 US-09-174-937-5 x HSU02570  ..

 Align seq 1/1  to: HSU02570  from: 1  to: 3345

      60 ArgIleLysLeuThrAlaProAsnIleAsnIle Ser   LeuAspGlnSe 75
         |||   :::|||:::...  ....:.  ......  |||......
      42 AGGGCAAGGCIGICIGGGACAGGCIIGGCCAI BATCCGCTCTCAGAGCT 91

      75 rGluGlySerIleLeuSerAspAspAsnLeuAspSerProAspGluIleA 92
          :::   ::::::     |||||:::  ......  ........
      92 GCAGGATGATCTGACCTTGGATGACACTAGCGACGCTCTGAACCAGCTGA 141

      92 spIle...AsnValAspGluLeuAspThrProAspGluAlaAspSerPhe 107
          :::   :::::::|||||   :::   |||
     142 AGCTGGCCTCCATCGATGAGAAGAACTGGCCC                    173

     108 GluTyrThrGlyHisAspProThrAlaAsnLysAspSerGlyGlnGluSe 124
                                                 ||
     174 ....                              ..............TC 175

     124 rGlSerIlProGlTyrThrAlaGlGlG ArgGluAspAsnArgL 141
          ::::::::::|'::::...   .           .
     176 GGATGAAATGCCTGACTTCCCCAAGTCAGATCA TCCAAAAGCAGC     221

     141 euTrpMetThrValValIleGlyGluGlnGluG nArgIleAspMetLys 157
                                :::   |||          :::|||
     222 .......................TCCCCGGAA TTGTCACACACCCTGAAG 248

     158 ValIleGluProTyrArgArgValIleSerHis               Gl 169
             :::|||||||   :::      |||          ::
     249 TGGGATGACCCATACTATGACATCGGCCGGCA AGATCGTGCAGGTGGC 298

     169 yGlyAspSerGlyTyrTyrGlyAspGlyLeuAs AlaIleIleValPheA 186
         :|||||:::   |||            :::   ||||| ||||||:
     299 AGCACATCACAAGTATGGG          C TAAGATCATTGTGTTTA 339

     186 laAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMet 202
         ||||||'   :::|||   |||   :::    '|:::
     337 GTGCCTGTCGAAATCCCCGCCAGCACCAGTTCAGCCAT ... ....... 374

     203 GluAsnLeuPheLeuTyrValIleSerThrLe GluLeuMetValAlaGl 219
         :::::: |:::  '||:::    '| || :::      |||   ::
     375 AGCAAGCTCCTGGGGTACCTGAAGCACACCCTGACCAGTACGTGGAGAG 424

     219 sAspTyrMetIleValTyrLeuAsnGlyAlaTh ProArgArgArgMetP 236
         :|||||   :::::::||||||:.  ...          ...   |
     425 TCATACACACTTTTGTACTTGCACACGCCC ACCAGCGACAACAAGT 474

     236 rGlyLeuGlyTrpMetLysLysCysTyrGlnM tIleAspArgArgLeu 252
         | :::| :::|'::::::   :::|'|:::   :::|||.|:::
     475 CCTCCCTCAGCTGGCTCGGTGATGCCTA CGGA AGTTTGACCGCAAGTAT 524

     253 ArgLysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleAr 269
         :::|:::|:::||:::::   |||||||| ||:::   ||||||::
     525 AAGAAGAACATCAAGGCCTTGTACATCGTGCA CAACCATGTTCATCAA 574

     269 rThrIleLeuAlaValThrArgProPheLeuS SerLysPheSerSerL 285
         :|||:::|||   :::   ....|||.. |||| ||||||  .   :|
     575 AACTCTGTCATCCTCTTCAAGCCCCTCATCA TTCAAGTTCGGGCAGA 624

     286 ysIleLysTyrValAsnSerLeuSerGluLeu erGlyLeuIleProMet 302
         .:::   |:| :|:|   ||||'|||. |  '          ...  ...
     625 AGACTTTTATATGAATTACTTGACGACTTA GAATACGTCAAGTTG 674

     303 AspCysIleHisIleProGluSerIleIleAs IleAspLeuLysLeuLy 319
         :::   :::   |||||.................   |||      |||||
     675 GAGCAGCTGGGGATCCCTGGCCAAGTGCTCAAA ATGACGACTTCCTGAA 724

     319 s 319
```

```
      !
    725 A 725

seq_name: gb_htg3:AC011488

seq_documentation_block:
LOCUS       AC011488   179301 bp    DNA             HTG        [illegible]99
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_17[illegible], *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *                   gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *                   gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *                   gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *                   gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *                   gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *                   gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *                   gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *                   gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *                   gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *                   gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *                   gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *                   gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *                   gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *                   gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *                   gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *                   gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *                   gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *                   gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *                   gap of unknown length
            *    15017    15614: contig of 598 bp in length
```

```
*                    gap of unknown length
*    15615    16687: contig of 1073 bp in length
*                    gap of unknown length
*    16688    18152: contig of 1465 bp in length
*                    gap of unknown length
*    18153    19221: contig of 1069 bp in length
*                    gap of unknown length
*    19222    20477: contig of 1256 bp in length
*                    gap of unknown length
*    20478    21388: contig of 911 bp in length
*                    gap of unknown length
*    21389    22059: contig of 670 bp in length
*                    gap of unknown length
*    22059    23154: contig of 1096 bp in length
*                    gap of unknown length
*    23155    23925: contig of 771 bp in length
*                    gap of unknown length
*    23926    24190: contig of 265 bp in length
*                    gap of unknown length
*    24291    24907: contig of 617 bp in length
*                    gap of unknown length
*    24909    25839: contig of 932 bp in length
*                    gap of unknown length
*    25840    27116: contig of 1277 bp in length
*                    gap of unknown length
*    27117    27904: contig of 788 bp in length
*                    gap of unknown length
*    27905    28519: contig of 615 bp in length
*                    gap of unknown length
*    28520    28930: contig of 411 bp in length
*                    gap of unknown length
*    28931    30384: contig of 1454 bp in length
*                    gap of unknown length
*    30385    31245: contig of 861 bp in length
*                    gap of unknown length
*    31246    32146: contig of 901 bp in length
*                    gap of unknown length
*    32149    33058: contig of 910 bp in length
*                    gap of unknown length
*    33059    34020: contig of 962 bp in length
*                    gap of unknown length
*    34021    34782: contig of 762 bp in length
*                    gap of unknown length
*    34783    36041: contig of 1259 bp in length
*                    gap of unknown length
*    36042    37033: contig of 992 bp in length
*                    gap of unknown length
*    37034    37565: contig of 532 bp in length
*                    gap of unknown length
*    37566    38344: contig of 779 bp in length
*                    gap of unknown length
*    38345    39359: contig of 1015 bp in length
*                    gap of unknown length
*    39360    41242: contig of 1883 bp in length
*                    gap of unknown length
*    41243    41839: contig of 597 bp in length
*                    gap of unknown length
*    41840    42836: contig of 997 bp in length
*                    gap of unknown length
*    42837    43913: contig of 1077 bp in length
*                    gap of unknown length
*    43914    55087: contig of 11174 bp in length
*                    gap of unknown length
*    55088    81867: contig of 26780 bp in length
*                    gap of unknown length
*    81868   112600: contig of 30733 bp in length
*                    gap of unknown length
*   112601   179301: contig of 66701 bp in length
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47123 c  47411 g  41894 t      [illegible]
ORIGIN


alignment_scores:
            Quality:  246.60              Length:     124
              Ratio:  2.004                 Gaps:       4
 Percent Similarity:  19.743   Percent Identity:  [illegible]

alignment_block:
 US-09-174-937-5 x AC012498/rev  ..

 Align seq 1/1  to reverse of: AC012498  from: 1  to: 174471

     115 ThrAlaAsnLysAspSerGlyGlnGluSerGlnSerIleProGlnTyrIk 141
         |*** ||||||       ||||||||||||||   |||   |||
  121984 ACGGCAAGAACATCGCCGGGAACTCCATCTCGCAGTCCCCTTAC 122028

     131 rAlaGluGluGlnArgGlnAspAs....ArgLeuIlpMetThrValValI 167
         |||| ||||                                  |||
  121934 CACCCAGGACGCCACCGCCGCCAACGCCGCTGTCGCGCAGTCTCACN 121984

     147 leGlyGluGlnGluGlnArgIleAspValLysValIleSlsProTyrAra 191
         ||||  |||  |  |||  |  |||  |||  |||
  121884 TCGGGGAGCAAGAGCACCGTATAGACGTGAACATGATCGGGCTTATAIG 121934

     164 ArgValIleSerHisGlyGlyAsp..................................... 171
         |||   |||||| '  ' ||||
  121834 AAAGTGGTCACCCACGGACGTGATAGCCCCCTTCCTGCCCCCCCCG 121878

     171 ..... 171

  121784 CTCCAGCCCGGCCCACTGGGCAAAAGGGGGTTTTCAGTTTTTTTTTG 121734

     171 ..... 171

  121734 ATGCACGGGGGATGTTAAGCCGTCAACTTTTTTTGGTGGATTTAATTTTG 121684

     171 ..... 171

  121684 GCAAGGCGTGCATGGTCAGGGACTTGTATTGGGGCTCCTGATGTTGGTT 121634

     171 ..... 171

  121634 GTCACTCCTCAGCGAAGCCACACTCCCTTTAATGGTGCTGTTTTTTTT 121588

     171 ..... 171

  121584 GGCTCGGAGCCATGCCCTCCCGAATCGTGTGGGTACCTTGATTTTTT 121538

     171 ..... 171

  121534 GGTGGGGGGCAGGCACCTCACCGCTTTTTTTGTGATTTTTATTTTTTTAT 121484

     172              SerGlyTyrTyrGlyAspGlyLeuAsnAlaLeuLeuV 214
                      |||||||||||   |||
  121484 GGGCTGCCTCTCCTCAGGGTACTACTGTTGAAGGCCTCAA  GATATATG 121434

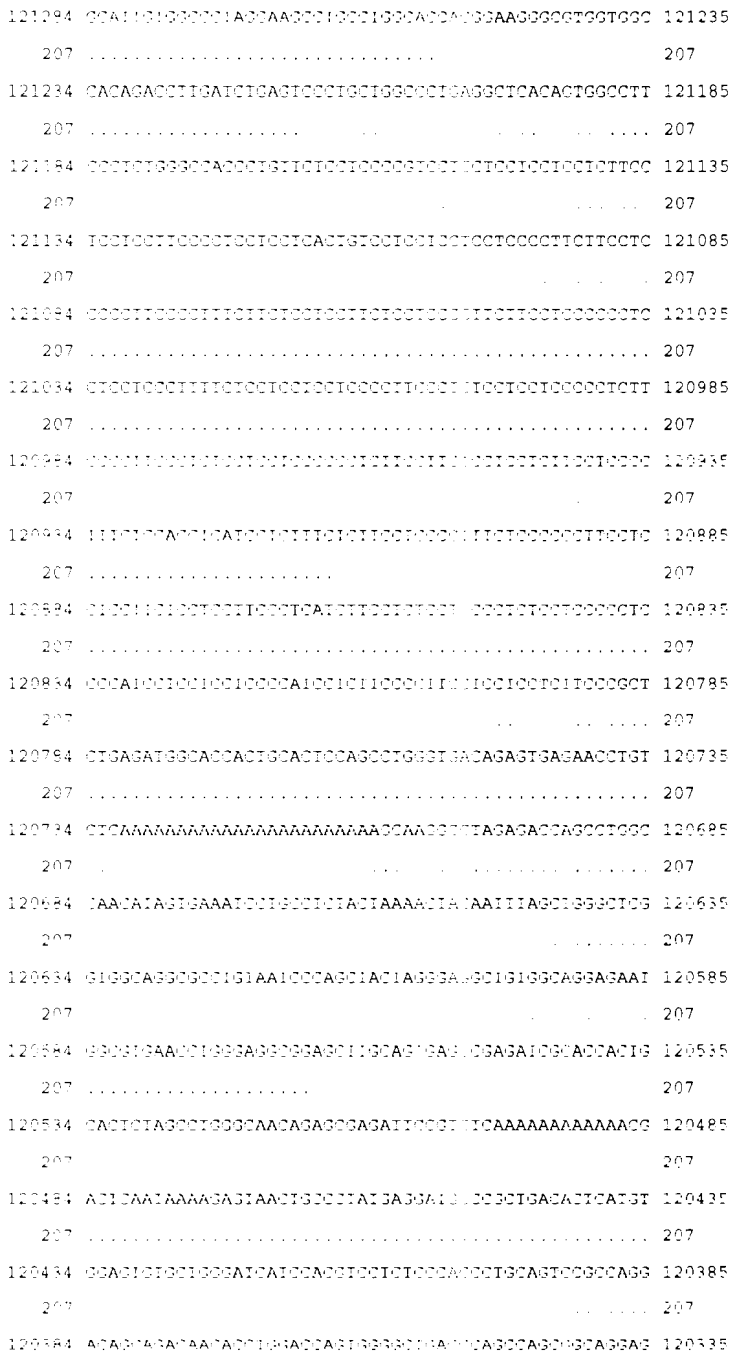
     184 alPheAlaAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyr 2
         |||| |||| |||| |||| |||| |||| |||
  121434 TCTTCGCAGGCTGCTTCCTTCCCAGACTACTCGCTCGACTATCATAT 121384

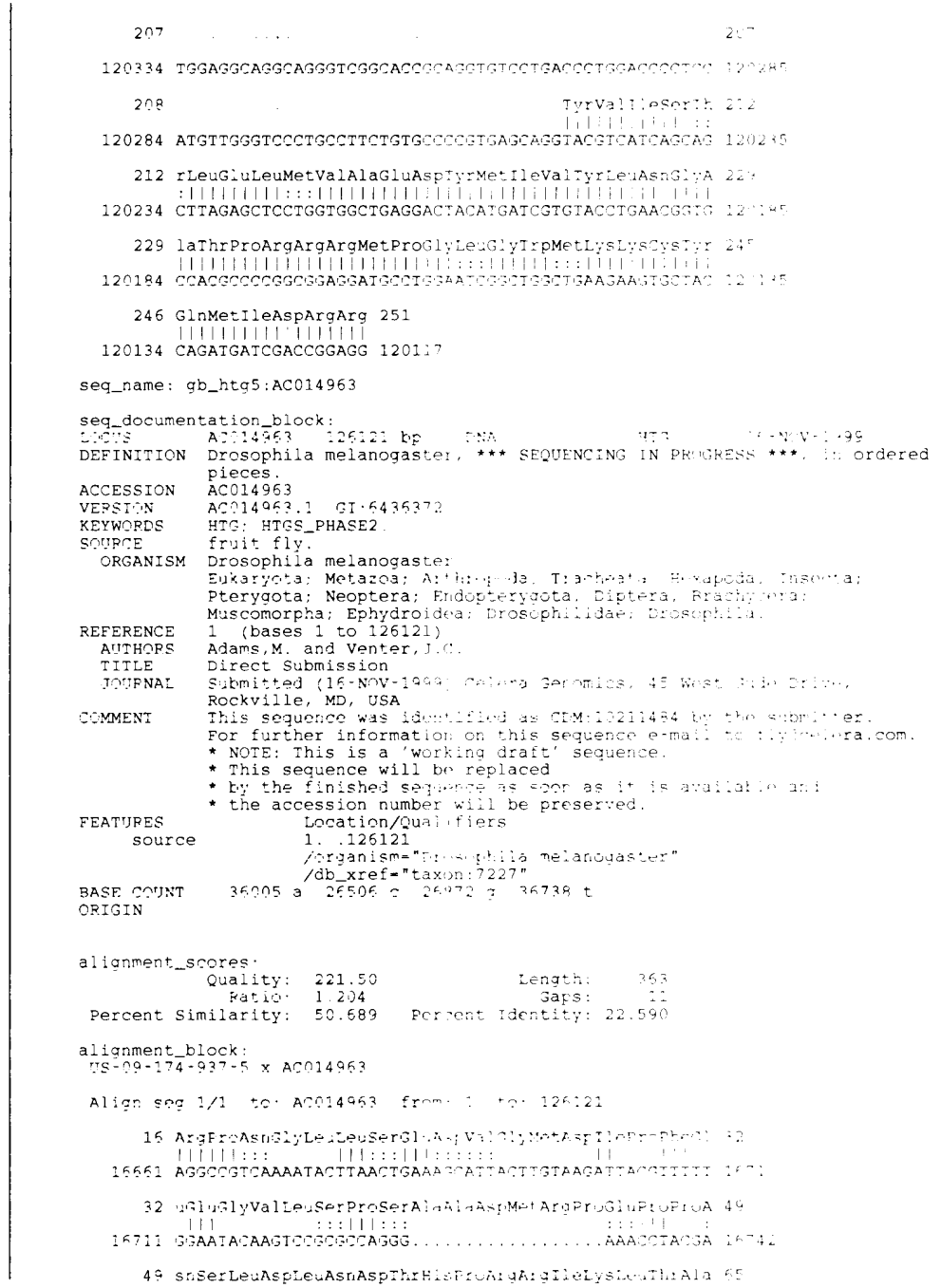
     201 ValMetGluAsnLeuPhe..Leu.............................. 2
         |||  ||| |||||||  |||
  121384 ATCATGGAGAACCTCTTCCTGTGAGTCCCTGTCGTGCGACTACTATAG 121334

     207                                 ..... 2

  121334 GGCCAGCTCTGATGCCTCCCTGCTCACAACGGGTACCAGCCTTTAATTTATT 121284

     207 ..... 2
```





```
seq_name: gb_htg5:AC014963

seq_documentation_block:
LOCUS       AC014963   126121 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, 18 ordered
            pieces.
ACCESSION   AC014963
VERSION     AC014963.1  GI:6436372
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 126121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211484 by the submitter.
            For further information on this sequence e-mail to trp@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    36005 a  26506 c  26972 g  36738 t
ORIGIN


alignment_scores:
           Quality:  221.50            Length:    363
             Ratio:  1.204               Gaps:     11
 Percent Similarity:  50.689   Percent Identity: 22.590

alignment_block:
 US-09-174-937-5 x AC014963

 Align seg 1/1  to: AC014963  from: 1  to: 126121

      16 ArgProAsnGlyLeuLeuSerGluAspValGlyMetAspIleProPheG 62
         |||||||::     |||:::|||::::::     ||    ::: :
   16661 AGGCCGTCAAAATACTTAACTGAAAGCATTACTTGTAAGATTACCTTTT 16711

      32 qGluGlyValLeuSerProSerAlaAlaAspMetArgProGluProProA 49
          |||       :::|||:::                  :::|| :
   16711 GGAATACAAGTCCGCGGCCAGGG.................AAACCTACGA 16742

      49 snSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAla 95
```

```
16743 GATTTATAAATGCCTTCCTCTCGCACCCCAAGGT.....  .....  16776

   66 ProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAs 82
16777 CCCGCCATCAATCCGATCGTC....                .. 16797

   82 pAspAsnLeuAspSerProAspGluIleAspIleAsnValAspGluLeuA 99
16798  GATAACTCGGATGAGCCCCAACCCAGTCTCTCCCATCTGCACGACTTCG 16846

   99 spThrProAspGluAlaAspSerPheGluTyrThrGlyHisAspProThr 115
16847 AGCCCAAGCTAGAGTTTGATGATACCGAGCTCCTGGCCCCATCTCCGCTT 16896

  116 AlaAsnLysAspSer                          .. 120
16847 GAAAGTAAATTGAATATTTAAAGAATCTACAAAATCATAGTCAAACTAAT 16946

  121 ...       GlyGlnGluSerGluSerIleProGluTyrThrAlaGluG 134
16947 TCTTCGATTTTTCCCAGAAGATGTTATGGTCGGTGATTTCGTTCTGGCCG 16996

  134 luGluArgGluAspAsnArgLeuTrpMetThrValValIleGlyGluGln 150
16997 AAGGTAAG..  ...  .....  TGGCGGAATCTACTTCTCCTGGCCAAT 17031

  151 GluGlnArgIleAspMetLysVal  .....IleGluProTyrArgAr 164
17032 GAATAACGTTTCGCATTATGCCTAGATCCAGAAGTTGAGCCCGAGGAAGA 17081

  164 gValIleSerHisGlyGlyAsp..                .... 171
17082 TGTAAATCCGCTGGAGGACGACTTCGAGGACCAACTCAGGGAGCAGTCTG 17131

  172                          SerGlyTyrTyrGly 176
17132 AAAACTTTCAGACTCCCAGAAACAAATGCGATTTCCTGGGAACCGATAAG 17181

  177 AspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSe 193
17182 CAGGGTCGCCACATCTTTGGCATTTATGCATCCCGCTTT...CCCGAAAA 17228

  193 rSerArg.          .....  ....  ....  .AlaAspTyrHisT 200
17229 ATCCCAACTTGAAGGATTTGTCCCGTATGCATTCATATTCGAGTACATTTT 17278

  200 yrValMetGluAsnLeuPheLeuTyr..        ..     .ValIleSer 211
17279 TTGTTCAATCATTAATCTATTTCCATTTGTCTATCAGGGAAATCATCAAG 17328

  212 ThrLeuGluLeuMetValAlaGluAspTyrMetIleValTyrLeuAsnGl 228
17329 GAGATCGAACCTTTTGTGGAGAACGACTACATTTTGGTTTACTTCCATCA 17378

  228 yAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLysCysI 245
17379 AGGCCTCAAGGAGGACAACAAACCATCCGCAGAGTTCCTGTGGAATTCGT 17428

  245 yrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSerPheIleIle 261
17429 ACAAGGAGCTCGATCGCAACTTCCGCAAGAAATCTGAAGACGTTGTATGTG 17478

  262 ValHisProSerTrpPheIleArgThrIleLeuAlaValThrArgProPh 278
17479 GTTCATCCGACGTGGTTCATACGGGTCATTTGGAACTTCTTTAGCCCGTT 17528

  278 eIleSerSerLysPheSerSerLysIleLysTyrValAsnSerLeuSerG 295
17529 CATTAGCGACAAGTTCCGCAAGAAGCTAGTCTACATCTCCAGCTTGGACG 17578

  295 luLeuSerGlyLeuIleProMetAspCysIleHisIleProGluSerIle 311
17579 AGTTGCGCCAGGCGCTCGGCCTGAATAACCTGAAGTGCGGACAACATC 17626

  312 IleAsnIleAspLeuLysLeu.....LysGluLysPro 322
17629 TGCGACCTGGACGACAAACTGAA.....GGGAAG CA 17667

seq_name: gb_pr2:HSU62794

seq_documentation_block:
LOCUS       HSU62794     1378 bp    mRNA            PRI       28-AUG-1997
DEFINITION  Human CDC42 GAP-related protein mRNA, partial cds.
ACCESSION   U62794
VERSION     U62794.1  GI:1916170
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 269 (50), 31859-31862 (1994)
  MEDLINE   95075275
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   3  (bases 1 to 1378)
  AUTHORS   Barfod,E.T. and Ashkenazi,A.
  TITLE     An effector function for Cdc42GAP that regulates the mitotic
            spindle
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Chan,M., Kuang,W.J. and Ashkenazi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1996) M.J. Geller Sr., 1 Bay Area Blvd., Mail Point
            San Bruno Blvd., S. San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CML"
                     /cell_type="eukaryocyte"
     CDS             <1. .672
                     /note="CDC42G-RP; similar to CDC42 GTPase-activating
                     protein encoded by GenBank Accession Number U02570;
                     insertion in the rhoGAP active site renders active site
                     boxes 2 and 3 out of frame and untranslated"
                     /codon_start=1
                     /product="CDC42 GAP-related protein"
                     /protein_id="AAB66594.1"
                     /db_xref="GI:1916171"
                     /translation="MNLLRSLSLLLLVSGAAGMGKLPSVAAVPPALLLSCPPCL
                     CPVPVNIVALYVVFTMFTVFLLTFVTFKFKGLVTPGVVLSLLSLFCMVKLEQLGI
                     PRIVLPVSGFLVSLVSFGTAMKPMPTPSTDFNLLTCVSLMLLCMNFLQEPIPIVLR
                     ETVALCAHALTTEGIPPRSANILVPLV..AVNKLSLPSSLAELLSPCPHPRAACG
                     FRPVQ"
     misc_feature    370. .402
                     /note="encodes SH3-binding domain; capable of binding src
                     and p85"
     misc_feature    465. .543
                     /note="active site box 1"
     misc_feature    585. .640
                     /note="55 nucleotide insertion into sequence identical
                     with CDC42GAP; insertion between active site boxes 1 and 2
                     shifts translational frame such that a stop codon occurs
                     at 670, and active site boxes 2 and 3 are not made"
     misc_feature    671. .763
```

```
                     /note="out-of-frame active site box 2"
     misc_feature    821. .949
                     /note="out-of-frame active site box 3"
     polyA_site      1378
                     /note="21 A nucleotides"
BASE COUNT      299 a    469 c    329 g    294 t
ORIGIN


alignment_scores:
            Quality:  157.00             Length:     69
              Ratio:  2.982                Gaps:      0
 Percent Similarity:  76.812   Percent Identity: 46.377

alignment_block:
 US-09-174-937-5 x HSU62794   ..

 Align seg 1/1  to: HSU62794  from: 1  to: 1378

     251 ArgLeuArgLysAsnLeuLysSerPheIleIleValHisProSerTrpPh 267
         :::  ::: :::::::::::::::::  :: :::::::::::  ::
     136 AGGTACAAGAAGAACATCAAGGCCTTGTACATCTGGCATCCAACCATGTT 185

     267 eIleArgThrIleLeuAlaValThrArgProPheIleSerSerLysPheS 284
         ::: ::::::::::::   :::   :::: ::::::::::   ::::::
     186 CATCAAAACTCTGCTCATCCTCTTCAAGCCCTCATCAGCTTCAAGTTCG 235

     284 erSerLysIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIle 300
         ::::::::::::   ::::::::::   :::::: ::::::::      :::
     236 GGCAGAAGATCTTCTATGTGAATTACCTGAGCGAGCTGAGCGAGCACGTG 285

     301 ProMetAspCysIleHisIleProGluSerIleIleAsnIleAspLeuLy 317
         ::::::   :::   ::::::: : : :  :::::   :::
     286 AAGCTGGAGCAGCTGGGGATCCCTCGCCAAGTGTTCAAATATGACGACTT 335

     317 sLeuLys 319
         ::::::
     336 CCTGAAA 342

seq_name: gb_htg1:CEY62F5

seq_documentation_block:
LOCUS       CEY62F5    310565 bp    DNA             HTG       03-OCT-1999
DEFINITION  Caenorhabditis elegans chromosome II clone Y62F5, *** SEQUENCING IN
            PROGRESS ***, 18 unordered pieces
ACCESSION   AL023633
VERSION     AL023633.27  GI:5123725
KEYWORDS    HTG; HTGS_PHASE1
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
REFERENCE   1  (bases 1 to 310565)
  AUTHORS   Williams,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 22, 1999 this sequence version replaced gi:4938509.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .310565
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y62F5"
BASE COUNT    94989 a  58692 c  57839 g  94951 t   4194 others
ORIGIN


alignment_scores:
            Quality:  146.50             Length:     327
              Ratio:  1.024                Gaps:      10
 Percent Similarity:  42.731   Percent Identity: 21.101

alignment_block:
 US-09-174-937-5 x CEY62F5   ..

 Align seg 1/1  to: CEY62F5  from: 1  to: 310565

      45 ProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIl 61
         |||:::|||   :::      ||'
   72822 CCGGATCCATCAGGTCCATCAGAT..........................  72847

      61 eLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlyS 78
                 ||||||||||||||||:::     ::::::     :
   72846 ............CCCAACATAAATCTTGGAGAATTCGAAGATCCGGTGG 72892

      78 erIleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsL 94
          :     ..!!!!!: :|||   :::|||||||:::::
   72883 ATGGCTATGCCGATGATGACCTTCTCACACCGATGACCTA......... 72923

      95 ValAspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThr 111

   72923 ..................      ..............  72923

     111 yHisAspProThrAlaAsnLysAspSerGlyGlnGluSerGluSer.... 126
                                 |||||:::   :::::::::
   72924 .....                 AGCGGTAAGTGGAATAATAATTATT 72944

     127 .. .. .. .. .. .. .. .. IleProGluTyrThrAlaGlyGlu 142
                               :::  :::  ::::::
   72949 GATATTTAATTTAATAGAAAACTGCTTTATTTCACTAAGTTTCACTTGAA 72998

     135 GluArgGluAspAsnArgLeuTrpMetThrValValIleGlyGluGlnGl 151
         |||.....:::|||  |||                    :::
   72999 TGACGCCATCTCGATCGTTTCTGGAA..............ACTAAAGA 73032

     151 uGlnArgIleAspMetLysValIleGluProTyrArg.............. 168
         !  :::   ::::::      :::|||||:::::
   73033 ATTCGAGACAGAGCTTGGAGGTGTTCAACCGTTCGAAGATCTTTTCAACG 73082

     164 .. .. .. .. .. .. .. .. .. ArgValIleSerHisGlyGlyAspSer 172
                               :::|||||||:::   |||
   73083 ACATATCAGCTCACGAAATTATTCAAGTAATTGCTGATGGGGATCGAGTT 73132

     173 GlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 189
         |||      ...:|||||:   :||  :::    :  |:::
   73133 GGA...AGGTAAATTAATTTGAATCCAAAAATTAATTATCTTTTTTGTTT 73182

     189 e.....  ..  .  . ............................... 189
         |
   73180 TAATGAATCATAGTTTGCAGGCCAACTGTTGTCGTATATGCTTATCGGCT 73229

     190 .. .. .. .. .. .. .. .. LeuProAspSerSerArgAlaAspTyr 198
                               :: :::| |          :
   73230 CCCATCAAGCAAGGAAATAGATCATCCAGACTTCTACAGTATTTGGTAC 73279

     199 HisTyrValMetGlu..... .....AsnLeuPheLeuTy 208
             :::::::::                 |||:::::::::
   73280 AAATTATTGATAAGGTTGGGAATTTTAATAAATATAATATTTTGNTAAA 73329

     208 rValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValT 225
         :::  :::::::  ::: ||   :::|||||   '||||||
   73330 AATAAAATTTGCAATTTTCCAGATCGTTGACCAAGACTACAAATTGTCT 73379
```

```
     225 yrLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMet 241
         ||.....  .  .....  ||  ...  ||...
   73380 ATTTTCACTATCGCCTACGAAGTCACAACAAACCACCGGTACGGTGGCTG 73429

     242 LysLysCysTyrGlnMetIleAspArgArg  leu  ..... 252
         ..... ||  ||||||||
   73430 TTCCAGGCTTACAAACAACTTGATCGTACCTAACTATTTGAATTCAATTC 73479

     253 ..  .  ..... ..ArgLysAsnLeuLy 257
                        ::::::::::::::
   73480 ATCTCTTCTCTTTCAAAAAAACCGGTTTCCACAATTCAAAAAGAACCTGAA 73529

     257 sSerPheIleIleValHisProSerTrpPheIleArgThrIleLeuAlaV 274
         |.....  .  ||||||  ||||||||  ||||||..
   73530 AGCGCTCTACGTAGTGCATCCAACACGATTCATCCGCATTATCTTCTCGC 73579

     274 alThrArgProPheIleSerSerLysPheSerSerLysIleLysTyrVal 290
         ::  :::  ||||||||||||||||||||||||||||  ||||||
   73580 TGTTCAAAGGTTTCATATCGAGCAAGTTTGAGAACAAGTTCCACTATGTC 73629

     291 AsnSerLeuSerGluLeuSerGlyLeuIle 309
         ... ||||||||||||  ...
   73630 ATGTGTATAGACGAGCTGGAGAACGCACTT 73659

seq_name. gb_in1.CEW02B12

seq_documentation_block:
LOCUS       CEW02B12    34305 bp    DNA             INV       12-NOV-1999
DEFINITION  Caenorhabditis elegans cosmid W02B12, complete sequence.
ACCESSION   Z66521
VERSION     Z66521.1  GI:1469136
KEYWORDS    HTG; Alanyl-tRNA synthetase; Esterase; Kinensin-like protein;
            Mitochondrial RNA splicing like protein; N-acetyllactosamine
            synthase; Phospholipase; Pre-mRNA splicing factor like protein;
            Serine/threonine protein kinase.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
REFERENCE   1  (bases 1 to 34305)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 34305)
  AUTHORS   Swinburne,J. and Ainscough,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1995) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jul 29, 1996 this sequence version replaced gi:1044853.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=W02B12
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the minimal insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we include 1-2 kb of
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone W02B12.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we usually include 1-2 kb of overlap between
            neighbouring submissions.
            The true right end of clone W02B12 is at 4161 in
            sequence Z66519.
            The true left end of clone B0024 is at 34218 in this sequence. The
            start of this sequence (1..1214) overlaps with the end of sequence
            AL031269.
            The end of this sequence (31218..34305) overlaps with the start of
            sequence Z66519.
FEATURES             Location/Qualifiers
     source          1..34305
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome "II"
                     /clone="W02B12"
     gene            4120..6453
                     /gene="W02B12.1"
     CDS             join(4120..4132,4535..4707,4755..4872,4942..5089,
                     5135..5301,5355..5518,5570..5711,5764..6154,6261..
                     6313..6453)
                     /gene="W02B12.1"
                     /note="similar to phospholipase; cDNA EST yk470h3.3 comes
                     from this gene; cDNA EST yk470h3.5 comes from this gene;
                     cDNA EST yk339e12.3 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA91399.1"
                     /db_xref="GI:1880427"
                     /db_xref="SPTREMBL:Q23114"
                     /translation="MPWLVTTLLQSLAMGLPLGLPNWGQLAFVMAPFKKVPISAHS
                     VPFADIKVIGALQDSLIAANQMCA,PGDDLAVLLVPSLAMLIDTKKSLDEHIIVANV
                     LRKFNFILVSASKIGSLNVWLVSLNNRLIGALKRLIIQLRALLYNMHAHSEINVK
                     EDWFLVNIFIAANSIVSEFYPLSTALNSNSTFRNIIDAYLIILNLFPRINFKF
                     FLFSPISVCPTFSPKFFVNSNIVSPTTVSTVMPNVMVPKTSPKVFFIEGLHTFEC
                     DCESNKQFTLSSILQVSFCNNAERELNTCSLSNKRLFTFVYLFTNSILEFPYASP
                     GVVDMTFFAPDCPHFSAVSHCNICDMLVNTIVLTVGFRLTSYNLSLFSVGLHCPSINC
                     PFFPTITKNSNNCASHFTLSELD"
     gene            complement(6752..7804)
                     /gene="W02B12.2"
     CDS             complement(join(6752..6847,6884..7191,7289..7315,
                     7605..7862,7915..8134))
                     /gene="W02B12.2"
                     /note="similar to pre-mRNA splicing factor like protein;
                     cDNA EST EMBL:D31383 comes from this gene; cDNA EST
                     EMBL:D65797 comes from this gene; cDNA EST EMBL:D72951
                     comes from this gene; cDNA EST EMBL:D69365 comes from this
                     gene; cDNA EST EMBL:D71420 comes from this gene; cDNA EST
                     EMBL:D70914 comes from this gene; cDNA EST EMBL:D74773
                     comes from this gene; cDNA EST EMBL:D74214 comes from this
                     gene; cDNA EST EMBL:D74214 comes from this gene; cDNA EST
                     EMBL:D74997 comes from this gene; cDNA EST EMBL:D79836
                     comes from this gene; cDNA EST EMBL:D29851 comes from this
                     gene; cDNA EST EMBL:C07345 comes from this gene; cDNA EST
                     EMBL:C09583 comes from this gene; cDNA EST yk389e5.3 comes
                     from this gene; cDNA EST yk389e5.5 comes from this gene;
                     cDNA EST yk334a4.3 comes from this gene; cDNA EST
                     yk334a4.5 comes from this gene; cDNA EST yk469g7.3 comes
                     from this gene"
                     /codon_start=1
                     /protein_id="CAA91394.1"
                     /db_xref="GI:1880428"
                     /db_xref="SPTREMBL:Q23112"
                     /translation="MVFVYIQRLPNRASDDDVEHIFRQYGKLSDVIMKNGFGFVDFQD
                     QRDALDAVHDLNGRELCGERVILEFPRPRKVSNEERRGSQFFSREPTFRGGERQFSN
                     RYSRPCSTRFFLVIENLSTRVSWQEIKDHFRKIGFPTVSDAHKRNVNLAIVCFTSHD
                     DLRPDAMNKLQGEDLNGRKLKCTDETRQRSRSRSPRRRSRSRSPTRSPSPPARRRSPGS
                     IFSIPRSMSASPRRSLSPASPESPSPSSPSPMSPFNPSLSPPSRDNSSPRS
```

```
                GSASP"
gene            9070..10722
                /gene="W02B12.3"
CDS             join(9070..9192,9478,9630..9802,10114..10187,
                10240..10320,10578..10722)
                /gene="W02B12.3"
                /note="similar to pre-mRNA splicing factor like protein;
                cDNA EST EMBL:D71552 comes from this gene; cDNA EST
                EMBL:D73263 comes from this gene; cDNA EST EMBL:D74207
                comes from this gene; cDNA EST EMBL:D76280 comes from this
                gene; cDNA EST EMBL:C08377 comes from this gene; cDNA EST
                EMBL:C10112 comes from this gene; cDNA EST EMBL:C07321
                comes from this gene; cDNA EST EMBL:C08652 comes from this
                gene; cDNA EST yk479g12.3 comes from this gene; cDNA EST
                yk479g12.5 comes from this gene; cDNA EST yk459f4.3 comes
                from this gene; cDNA EST yk459f4.5 comes from this gene;
                cDNA EST yk454f4.3 comes from this gene; cDNA EST
                yk454f4.5 comes from this gene; cDNA EST yk448b3.3 comes
                from this gene; cDNA EST yk448b3.5 comes from this gene;
                cDNA EST yk401b2.3 comes from this gene; cDNA EST
                yk401b2.5 comes from this gene; cDNA EST yk310g11.3 comes
                from this gene; cDNA EST yk310g11.5 comes from this gene;
                cDNA EST yk303a9.3 comes from this gene; cDNA EST
                yk303a9.5 comes from this gene; cDNA EST yk300f1.3 comes
                from this gene; cDNA EST yk300f1.5 comes from this gene;
                cDNA EST yk282b11.3 comes from this gene; cDNA EST
                yk282b11.5 comes from this gene; cDNA EST yk246c3.3 comes
                from this gene; cDNA EST yk246c3.5 comes from this gene;
                cDNA EST yk229b9.3 comes from this gene; cDNA EST
                yk229b9.5 comes from this gene; cDNA EST yk507e10.3 comes
                from this gene; cDNA EST yk519e3.3 comes from this gene;
                cDNA EST yk536b6.3 comes from this gene; cDNA EST
                yk570b9.3 comes from this gene; cDNA EST yk570c9.3 comes
                from this gene; cDNA EST yk583c6.3 comes from this gene;
                cDNA EST yk590f10.3 comes from this gene; cDNA EST
                yk595a1.3 comes from this gene; cDNA EST yk606c1.3 comes
                from this gene; cDNA EST yk613h10.3 comes from this gene;
                cDNA EST yk617d3.3 comes from this gene; cDNA EST
                yk637g10.3 comes from this gene; cDNA EST yk642h10.3 comes
                from this gene; cDNA EST yk657h1.3 comes from this gene;
                cDNA EST yk667a2.3 comes from this gene; cDNA EST
                yk675e1.3 comes from this gene; cDNA EST yk381h2.5 comes
                from this gene"
                /codon_start=1
                /protein_id="CAA91395.1"
                /db_xref="GI:3880429"
                /db_xref="SPTREMBL:Q23121"
                /translation="MAARIYIGRLTSPVSEKDIEHFFPGYGQIRDVLLKNGFGFVEFD
                DKRDAEDAVHDLNGKELGGERVILDYSKPRGGGGDRGGFGGGGRGGARVSSYSGGGGG
                GPDRFDRYDRGPPPRPESPYGRPYSTRHPVVVENLSSPISWQDLKDQVRRQGVEPTYAE
                AHKRPNEALLCFATPSDLKRCIEKCDGMDLNGRKIKMIDDSQAGPSFSPSNSRSRSRS
                RSPDRPPSPSPSSSPSKSPSRSPPKPSPPESKSKSRSRSPSRSADNPKSPSPSPSKK
                VDRSPSPPRGSRSPSEKGSPRRSRSASPMDNGDGDN"
gene            complement(11094..14445)
                /gene="W02B12.4"
CDS             complement(join(11094..11199,11252..11517,11739..11804,
                11859..12049,12121..12337,12380..12510,12572..13073,
                13651..13704,13748..13894,14236..14445))
                /gene="W02B12.4"
                /note="similar to esterase"
                /codon_start=1
                /protein_id="CAA91397.1"
                /db_xref="GI:3880431"
                /db_xref="SPTREMBL:Q23123"
                /translation="MNWVQTCFLLIFLTTLVFFGTADPLETSADSPVLSTGLGYIPGP
                RVSTKSGFLVDYFTGVPFAEKPEPLKKSKLVNPWVDIRNTQQYPPKCIPAARPVNISE
                FSEDCLYMNILKPSSLNITPHYFVVVFVHGGGFQEVLFFSILNVLGDGNDASDAFAE
                HFVSKEMIVVTIQYRLGMLGFLSLGCSDPLIPGNLGLWDMVTAFRFINATIENFNGDI
                NNMTLMGHSAGAMAVSLHSMSPISSVHFKKYIQLSSSSWGLTRYKDQNIVQTKKILKD
                LNCDISSAKSIVDCVNSTSLDSMYSAQVCAHPFSFEYIFSRVETGYITFAPKIANFPI
                CVIPLFVQEILKTRVDTTLWPEFGDELIPDIPDNMVKNTQNQKVLMGVMTLESLYFSE
                FRFFVNCQIYKKNVGFVNDLKQCTSYFPAYLKSSPAIFPYIHPDTVEEYINSTIAIQY
                GIGSESAYDMTSAYYLSANATDNDYGYFMWQRSKVSARQMFLKSTVTLVALEFPKKID
                SNTGFDIPILREARARSQVSPDLYLYHFPYFNPAEFDVDFFVRATYHCHEFPYIWGVY
                KDNYFEPNDDDPTVSAFLITSLTNFIKTGNPSSSFPNWPLYDQYSMTYIGPKPTTDK
                RLLTL"
gene            complement(15419..19110)
                /gene="W02B12.5"
CDS             complement(join(15419..15430,15540..15729,16310..16580,
                16632..16884,16992..17155,17204..17299,17446..17868,
                17926..18642,18696..18986,19033..19110))
                /gene="W02B12.5"

alignment_scores:
            Quality:  146.50              Length:    327
              Ratio:  1.024                 Gaps:     10
 Percent Similarity:  43.731   Percent Identity: 21.101

alignment_block:
 US-09-174-937-5 x CEW02B12  ..

 Align seg 1/1  to: CEW02B12  from: 1  to: 34305

      45 ProGluProProAsnSerLeuAspLeuAsnAspThrHisProArgArgIl 61
         |||:::|||   :::     |||
   20004 CCGGATCCATCAGGTCCATCAGAT ..... ..... 20027

      61 eLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlyS 78
                 ||||||||||||||||:::     :::::::    :
   20028 .....CCCAACATAAATCTTCGACAATTCGAAGATCCGGTGG 20064

      78 erIleLeuSerAspAspAsnLeuAspSerProAspGluIleAspIleAsn 94
         :      ::::|||||||:::|||   :::|||||||:::::
   20065 ATGGCTATGCCGATGATGACCTTCTCACACCGGATGACCTA 20105

      95 ValAspGluLeuAspThrProAspGluAlaAspSerPheGluTyrThrGl 111
   20105 ..... ..... ..... ..... 20105

     111 yHisAspProThrAlaAsnLysAspSerGlyGlnGluSerGluSer.... 126
                                 ||||||||:::  :::::::::
   20106 .....ASCGGTAAGTGGAATAATAATTATT 20130

     127 .  ..... .....  IleProGluTyrThrAlaGluGlu 134
                        :::  :::  ::::::  |||
   20131 GATATTTAATTTAATAGAAAACTGGIIIAIITCAGTAAGIGICACTGGAA 20180

     135 GluArgGluAspAsnArgLeuTrpMetThrValValIleGlyGluGlnGl 151
         |||:::  :::|||:::|||                    :::||
   20181 TGACGCCATCTCGATCGTTTCTGCAA.  . . . ACTCAACA 20214

     151 uGlnArgIleAspMetLysValIleGluProTyrArg..... 162
         |  :::  ::::::     :::||||||||:::::
   20215 ATTCGAGACAGAGCTTGGAGGTGTTGAACCGTTCGAAGATCTTTTCAACC 20264

     164 .....ArgValIleSerHisGlyGlyAspSer 172
                    :::|'|||||:::  |||
   20265 ACATATCAGCTCACGAAATTATTCAAGTAATTGCTGATGGGGATCGAGTT 20314

     173 GlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 189
         |||          :::|||||||     |||  :::     |||||
   20315 GGA . AGGTAAATTAATTTGAATCCAAAAATTAATTATCTTTTTGTTT 20361

     189 e..... 189
         |
   20362 TAATGAATCATAGTTTGCAGGCCAATTGTTGTCGTATATGCTTATCGGCT 20411

     190 . . . ..... . ..LeuProAspSerSerArgAlaAspTyr 198
                       :::|||||                 :::
   20412 CCCATCAAGCAAGGAAATAGATCATGCCAGACTTCTACAGTATTTCGTAC 20461

     199 HisTyrValMetGlu.....          .AsnLeuPheLeuTy 208
         :::::::::          |||:::::::::::
   20462 AAATTATTGATAAGGTTGGGAATTTTTAATAAATATAATATTTTGATAAA 20511

     209 rValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValT 225
```

```
          :::      :.....     ...|||   ...|||||||   |||||||
  20512 AATAAAATTTGCAATTTTCCAGATCGTTGACCAAGACTACACAATTGTCT 20561

    225 yrLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMet 241
        ||:.....  :...      :.....    ||   ...     |||...
  20562 ATTTTCACTATGTTTATAGAAATTCACAACAAAATCATCGGTACGGTGGTG 20611

    242 LysLysCysTyrGlnMetIleAspArgArg..Leu..............  252
        :::::..||.........:|||||||||
  20612 TTCCAGCGTTACAAACAACTTGATCGTAGGTAAGTATTTGAATTCAATTC 20661

    252                    .       .....          .ArgLysAsnLeuLy 257
                                                  :::|||||||||||
  20662 ATCTCTTTCTCTTTCAAAAAAACCGGTTTCCAGATTCAAAAAGAACCTGAA 20711

    257 sSerPheIleIleValHisProSerTrpPheIleArgThrIleLeuAlaV 274
        |:::::.   :::|||||||||||::   |||||||||   |||:::.:.
  20712 AGCGCTTCTACTTAGTGCATCCAACAGGATTCATCCGCATTATCTTCTCGC 20761

    274 alThrArgProPheIleSerSerLysPheSerSerLysIleLysTyrVal 290
        ::   :::  |||||||||||||||||:::::|||:::   ||||||
  20762 TGTTCAAAGGTTTCATATCGAGCAAGTTTGAGAACAAGTTCCACTATGTC 20811

    291 AsnSerLeuSerGluLeuSerGlyLeuIle 300
            :::::::|||||:::   .
  20812 ATGTGTATAGACGAGCTGGACAACGCACTT 20841

seq_name: gb_in2:AC004283

seq_documentation_block:
LOCUS       AC004283     80095 bp    DNA             INV       29-AUG-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1 DS05557 (D152)), complete
            sequence.
ACCESSION   AC004283 AC002945 AC002947 AC003240 AC002948 AC003241 AC002949
            AC002950 AC003242 AC002951 AC002952
VERSION     AC004283.1  GI:3492867
KEYWORDS    HTG
SOURCE      Drosophila melanogaster (Subclones in sac from P1 clone DS05557
            (D152)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 80095)
  AUTHORS   Celniker,S.E., George,R.A., Galle,P.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2P, region 43E8-43E12
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 80095)
  AUTHORS   Celniker,S.E., George,R.A., Galle,P., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Aug 29, 1998 this sequence version replaced gi:3419828.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://[illegible]) or send
            email to bdgp@fruitfly.lbl.gov.
            Library location: 85-88.
FEATURES             Location/Qualifiers
     source          1..80095
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="43E8-43E12"
                     /clone="P1 DS05557 (D152)"
BASE COUNT    23315 a  17568 c  16970 g  22242 t
ORIGIN


alignment_scores:
            Quality:  139.00              Length:    242
              Ratio:  1.904                 Gaps:      1
 Percent Similarity:  55.303   Percent Identity:  25.0[illegible]

alignment_block:
 US-09-174-937-5 x AC004283  ..

 Align seg 1/1  to: AC004283  from: 1  to: 80095

    207 LeuTyrValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIl 223
        ||||||...|||.||   | ||||   :::                 :::
  39383 CTATATCTGATTAAGCTACTGGATCCCATCGTTAAAGGAGACTATTTCAT 39432

    223 eValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyT 24
        |   |||.....:      |'    ::::::   :  :::
  39433 CTCGTACTTCCACACGCTGACATTCCACGAATAACTATCCATTCCTCCATT 39482

    240 rpMetLysLysCysTyrGlnMetIle..................... .... 249
        ||:::::::::   ||  ::::::::
  39483 GGCTGCGTGAGGTCTACAGTGTGTTGCCTTACAAGTTAATTTTTATCTTTA 39532

    249 ...      .  .....                        .....AspArgAr 261
                                                      :::
  39533 ACACCTCATGGAAAATGACATTTTAATAAATTTGCTTATGTTCCGTTAG 39582

    261 gLeuArgLysAsnLeuLysSerPheIleIleValHisProSerTrpPhel 268
        |  ...|||||||||||:::'  '  '  '  '  '  '  ::::::::
  39583 ATACAAGAAGAACCTGAAGTCTTTATATTGTCATCTTATTTTTGGA 39632

    268 leArgThrIleLeuAlaValThrArgProPheIleSerSerLysPheSer 284
        ..  ...  .:          :::::::      :::::::
  39633 CTAAGATGATTGATTTGTTCATCAATGTTATGTTGTTTTTAAAG 39682

    285 SerLysIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIlePr 301
        ....| |...      ...   ::::::::::::          :::
  39683 GCCAAGGTGCACTCGCTTCTGGGTGTGGAGTACTTGTACTCGCCATCAC 39732

    301 cMetAspCysIleHisIleProGluSerIleIleAsnIleAspLeu 316
           |||   ::::::'|'|'|         :   :::   :::
  39733 CAAGGATCAGCTGGAAATTCCCGTCTACATACTGATACGACATA 39778
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

QM nucleic nucleic search, using sw model

Run on: February 7, 2000, 19:33:28 ; Search time 3770.17 Seconds
(without alignments)
378.551 Million cell updates/sec

Title: US-09-174-937-9
Perfect score: 378
Sequence: 1 atgctgaagagctgtagtcg.....taagttggcaaggccaaagt 378

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score or a lower score if fewer were printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	271.2	71.7	308	24	[illegible]	[illegible] HUM;18G12B
2	259.4	68.4	275	94	AA[illegible]	AA[illegible] ab40a09.s
3	179.4	47.5	296	81	AA[illegible]	AA[illegible] EST10971
c 4	68.8	18.2	565	104	A[illegible]	A[illegible] RPCI-11-3
5	53.8	14.2	604	51	AI72214	AI72214, th19a03.y
c 6	46.2	12.2	572	79	[illegible]	AA[illegible] Furubripe
7	39	10.3	858	69	AI[illegible]	AI[illegible] AI07928l

	8	35	9.3	631	74	AW217381	AW217381 EST296065
c	9	34.8	9.2	249	20	T11739	T11739 A1147F Hear
	10	34.2	9.0	411	39	AQ178663	AQ178663 HS_2242_A
	11	34	9.0	468	35	Z99377	Z99377 HSZ99377 DK
	12	33.2	8.8	484	22	H11905	H11905 ym10c10.r1
	13	33.2	8.8	469	23	H18541	H18541 ym43b03.r1
c	14	33.2	8.8	469	46	AI417625	AI417625 tg80c11.x
c	15	33.2	8.8	402	69	AW104968	AW104968 xd73d06.x
	16	32.8	8.7	386	90	AQ001306	AQ001306 CIT-HSP-2
c	17	32.8	8.7	292	91	AQ095898	AQ095898 HS_3033_A
	18	32.6	8.6	720	45	AV003693	AV003693
	19	32.6	8.6	661	64	AW032533	AW032533 EST273789
c	20	32.6	8.6	466	104	AQ517821	AQ517821 HS_5133_B
	21	32.2	8.5	799	89	AQ897368	AQ897368 HS_3134_A
c	22	32	8.5	247	70	AV234505	AV234505 AV234505
	23	32	8.5	709	79	AJ009368	AJ009368 Homo sapi
	24	31.8	8.4	679	28	AA062958	AA062958 ze24e07.r
c	25	31.8	8.4	561	40	AA914235	AA914235 wy99f10.r
c	26	31.8	8.4	378	46	AI350920	AI350920 qg22g08.x
	27	31.8	8.4	512	59	AI980693	AI980693 pat.pk001
	28	31.8	8.4	514	74	AW186065	AW186065 se63b06.y
	29	31.6	8.4	454	26	W79207	W79207 zd79g04.r1
	30	31.6	8.4	536	33	AA416376	AA416376 vd18h11.s
c	31	31.6	8.4	354	41	AI034967	AI034967 LmLv33p10
c	32	31.6	8.4	508	64	AW019691	AW019691 fd55g12.x
c	33	31.6	8.4	747	81	B00855	B00855 cSRL-122a22
c	34	31.6	8.4	567	90	AQ014362	AQ014362 RPCI11-23
c	35	31.6	8.4	400	93	AQ072453	AQ072453 HS_3024_A
	36	31.6	8.4	517	99	AQ173633	AQ173633 HS_3204_B
	37	31.4	8.3	267	20	Z25841	Z25841 GGEST5227 C
c	38	31.4	8.3	316	21	T75530	T75530 yd63d07.r1
	39	31.4	8.3	350	47	AI466079	AI466079 vw39d07.y
	40	31.4	8.3	569	87	AQ783080	AQ783080 HS_5412_A
c	41	31.2	8.3	593	27	AA046941	AA046941 zf13g12.s
c	42	31.2	8.3	398	36	AA620103	AA620103 wl53h05.r
c	43	31.2	8.3	630	42	AI115241	AI115241 qi43a02.y
c	44	31.2	8.3	605	42	AI115255	AI115255 qi43b12.y
c	45	31.2	8.3	1191	79	CNS00059	AL060259 Drosophil

ALIGNMENTS

RESULT 1
D61567
LOCUS D61567 302 bp mRNA EST 14-DEC-1995
DEFINITION HUM419G12B Clontech human fetal brain polyA+ mRNA (#6535) Homo sapiens cDNA clone GEN-419G12 5', mRNA sequence
ACCESSION D61567
VERSION D61567.1 GI:970692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 302)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.
TITLE Fujiwara et al. (1995)
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1995 this sequence version replaced gi:795691.
Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd.
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-85-2555
Fax: 0886-37-1035
Insert Length: 1881 Std Error: 0.00
High quality sequence stop: 311.
FEATURES Location/Qualifiers
source 1..302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-419G12"
/clone_lib="Clontech human fetal brain polyA+ mRNA (#6535)"
BASE COUNT 79 a 78 c 65 g 79 t 1 others
ORIGIN

Query Match 91.7% Score 271.2 DB 24: Length 5 2:
Best Local Similarity 99.6% Pred. No. 1.9e-74:
Matches 271: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

Qy 1 atgctgaagagctgtagtcgtgcatttctcaccatcgtttgaaaggcctcctcttcat 60
Db 31 ATGCTGAAGAGCTGTAGTCGTGCATTTCTCACCATCGTTTGAAAGGCCTCCTCTTCAT 90

Qy 61 ctcagaagactatgtcagaggatgtaggaatggacatcccctttgaaggaggcctttta 120
Db 91 CTCAGAAGACTATGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGGAGGCCTTTTA 150

Qy 121 agtcccagtgctgcagacatgaggt ttcccctgctaattctctgtgt ttcttaatcta 180
Db 151 AGTCCCAGTGCTGCAGACATGA CTTCCCCTGCTAATTCTCTGTGTTCTTAATCTA 210

Qy 181 catcctcggagaattcaagctcatatcccaaatatcaatttttctctgcaaaaaattta 240
Db 211 CATCCTCGGAGAATTCAAGCTCA CCCAAATATCAATTTTTCTCTGCAAAAAATTTA 270

Qy 241 ggatctatttctctgatgataactttgaacag 272
Db 271 GGATCTATTCTCTCTGATGATAACTTGGACAG 302

RESULT 2
AA504063
LOCUS AA504063 275 bp mRNA EST 20-AUG-1997
DEFINITION nh40g09.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954880, mRNA sequence.
ACCESSION AA504063
VERSION AA504063.1 GI:2239030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 275)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:[illegible].
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rourik A. Armaqui, M.D., Michael P. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 404 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 219.
FEATURES Location/Qualifiers
source 1..275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954880"
/clone_lib="NCI_CGAP_Pr5"

/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from normal prostatic epithelial cells, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. "
BASE COUNT 60 a 66 c 62 g 67 t
ORIGIN

Query Match 68.4%; Score 278.4; DB 34; Length 275;
Best Local Similarity 99.6%; Pred. No. 1.8e-70;
Matches 259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 68 gactactgtcagaggatgtaggaatggacatccccttgaagagggcgtgctgagtccca 127
Db 16 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTGAAGAGGGCGTGCTGAGTCCCA 75

Qy 128 gtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcctc 187
Db 76 GTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTC 135

Qy 188 ggagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatcta 247
Db 136 GGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 195

Qy 248 ttctctctgatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttg 307
Db 196 TTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTG 255

Qy 308 ataccccccgatgaagcagat 327
Db 256 ATAGCCCCCGATGAAGCAGAT 275

RESULT 3
AA296475
LOCUS AA296475 236 bp mRNA EST 18-APR-1997
DEFINITION EST10971 Umbilical vein endothelial cells II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA296475
VERSION AA296475.1 GI:1948809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 236)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., , Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On Apr 14, 1993 this sequence version replaced gi:693143.
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1..236
/organism="Homo sapiens"
/db_xref="ATCC (human):1944265"
/db_xref="taxon:9606"
/clone_lib="Umbilical vein endothelial cells II"
/cell_type="endothelial cell"
/dev_stage="fetus"
/note="Organ: umbilical vein; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 66 a 54 c 57 g 55 t 4 others
ORIGIN

Query Match 47.5%; Score 173.4; DB 34; Length 275;
Best Local Similarity 97.8%; Pred. No. 8.9e-41;
Matches 180; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 68 gactactgtcagaggatgtaggaatggacatccccttgaagagggcgtgctgagtccca 127
Db 53 GACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTGAAGAGGGCGTGCTGAGTCCCA 112

Qy 128 gtgctgcagacatgaggcctgaacctcctaattctctggatcttaatgacactcatcctc 187
Db 113 GTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTC 172

Qy 188 ggagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatcta 247
Db 173 GGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTA 232

Qy 248 ttct 251
Db 233 TTCT 236

RESULT 4
AQ533091/c
LOCUS AQ533091 565 bp DNA GSS 14-MAY-1999
DEFINITION RPCI-11-351B1.TV RPCI 11 Homo sapiens genomic clone RPCI-11-351B1, genomic survey sequence.
ACCESSION AQ533091
VERSION AQ533091.1 GI:4844791
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 565)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.chori.org/) or from

```
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..565
                     /organism="Homo sapiens"
                     /db_xref="GDB:763442"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-351B1"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      142 a    150 c    114 g    159 t
ORIGIN


  Query Match             18.2%;  Score 69.8;  DB 104;  Length 565;
  Best Local Similarity   86.4%;  Pred. No. 4.7e-11;
  Matches   76;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 atgctgaagagctgtagtcgtgcgatcctttctgaccctccgttagaaagcctcctctcatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 ATGCTGAAGAGCTGTAGTCGTGCGATCCTTTCTGACCCTCCGTTAGAAAGCCTCCTCTCATC 133

Qy      61 ctcagaaagactctgtccagaggatgtag 89
           |||||||     ||      | |||||
Db     192 CTCAGAAAGGTAGGTGGGTCCTGGTGTAG 105


RESULT   5
AI722140
LOCUS       AI722140       604 bp    mRNA            EST       10-JUN-1999
DEFINITION  fd19g03.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2352052 5' similar to TR:Q15073 Q15073 KIAA0367 ;, mRNA sequence.
ACCESSION   AI722140
VERSION     AI722140.1  GI:5040469
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B.,
            Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1999 this sequence version replaced gi:2948370.
            Other_FSTs: fd19g03.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2352052"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACATGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [AT3733CCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site: CACATGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT      194 a    145 c    103 g    121 t      1 others
ORIGIN


  Query Match             14.2%;  Score 53.8;  DB 51;  Length 604;
  Best Local Similarity   66.0%;  Pred. No. 2.4e-06;
  Matches  101;  Conservative    0;  Mismatches   47;  Indels   15;  Gaps    1;

Qy     187 tcctcggagaatcaaagctcacagcccaaatatcaatcttttctctgga-aaaattaa-- 242
           | ||| ||| |  || ||  | || ||  ||| | |||| |  ||||||| |||  --
Db     239 TTCTCAGAGGAAAAAACTGTCC----AGTATCTTCAATGCAG----AAA---GA-- 285

Qy     243 atctattctctctgatgataacttaaaaatt.agatgaaatttaaat.a.tttgataa 302
                   |||||   ||.|||  |.|||||   |||| |
Db     286 ----------GGATGACCTGTTTCACAGGCAGATGACCTCGACATCAAATTGGATAA 343

Qy     303 acttgataccccccgatgaagcagattcttttga 335
           || || || || |||||||| | | |   |
Db     344 TCTGGACACACCTGATGAAGCAGATTATACAGA 376


RESULT   6
FR0009132/c
LOCUS       FR0009132      572 bp    DNA             GSS       14-SEP-1997
DEFINITION  F.rubripes GSS sequence, clone 010H20aD1, genomic survey sequence.
ACCESSION   AL000418
VERSION     AL000418.1  GI:2438270
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes, Tetraodontoidei, Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Moot,S., Warburton,P., Taylor,K.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB, Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid clone with
            sequence.
FEATURES             Location/Qualifiers
     source          1..572
                     /organism="Fugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 010H20"
                     /clone="010H20aD1"
BASE COUNT      108 a    142 c    141 g    147 t     34 others
ORIGIN


  Query Match             12.2%;  Score 46.2;  DB 7;  Length 572;
  Best Local Similarity   66.7%;  Pred. No. 0.00055;
  Matches   66;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    0;
```

```
Qy      256 gatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgatacccca 315
            || ||    |  ||| |  |||||| ||  | ||||||||| |||||||| ||||| || ||
Db      347 GACGACCTCCGGGGAGACGCCAGAGGACGCCGACATCAACGTGGATGAGCTTGACACGCCC 289

Qy      316 gatgaagcagattcttttgagtacactggccatggtaag 354
            || || |  || |   | || |||||||| ||| |  | |
Db      287 GACGAGAGGGACTACCTGGACTACACTGACGACGAGACG 249


RESULT   7
AU079281
LOCUS       AU079281      858 bp    mRNA            EST       20-OCT-1999
DEFINITION  AU079281 Sugano mouse brain mncb Mus musculus cDNA clone MNCb-4083
            5', mRNA sequence.
ACCESSION   AU079281
VERSION     AU079281.1  GI:6084036
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Hashimoto,K., Kusuda,J., Toyoda,A., Tanuma,R., Ito,A., Hirata,M.,
            Suzuki,Y., Sasaki,M. and Sugano,S.
  TITLE     Isolation of full-length cDNA clones from a mouse brain cDNA
            library made by oligo-capping method
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 22, 1999 this sequence version replaced gi:3247313.
            Contact: Katsuyuki Hashimoto
            Division of Genetic Resources
            National Institute of Infectious Diseases
            23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
            Email: khashi@nih.go.jp
            URL: http://www.nih.go.jp/yoken/genbank/.
FEATURES             Location/Qualifiers
     source          1..858
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="MNCb-4083"
                     /clone_lib="Sugano mouse brain mncb"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="TOP10"
                     /note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [AIGTGGCCTTTTTTTTTTTTTTT] double-stranded cDNA was
                     ligated to a DraIII adaptor [TGTTGGCCTACTGG] , digested
                     and cloned into distinct DraIII sites of the pME18S-FL3.
                     XhoI sites just outside the DraIII sites can be used to
                     isolate the cDNA insert.  Size selection was performed to
                     exclude fragments <1.5 kb. Library was constructed by
                     Sugano et al. (University of Tokyo, Institute of Medical
                     Science).  Custom primer for sequencing: 5' end primer
                     [CTTCTGCTCTAAAAGCTGCG]"
BASE COUNT      229 a    191 c    171 g    191 t     86 others
ORIGIN


  Query Match           10.3%;  Score 39;  DB 69;  Length 858;
  Best Local Similarity   50.5%;  Pred. No. 0.11;
  Matches   97;  Conservative    0;  Mismatches   92;  Indels    3;  Gaps    1;

Qy      145 ctgaacctcctaattctctggatcttaatgacactcatcctcggagaatcaagctcacag 205
            || ||| ||||    |  || | | |   |  |||      | ||  |   |   |   |
Db      509 CTTAACTTCCTTCAACCTTGAANTTGACCGGANCCCATNAAAANAAAAAAACCTTGGNGG 568

Qy      206 ccccaaatatcaatctttctctggaccaaagtgaaggatctattctctctgatgataact 265
              ||||| || ||  |||| || |||||    | || | | |  | ||  |  ||   |
Db      569 TTCCAAAAATAAAATTTTCCTTGAACCAAANCNAGGGNTTTTTGTTGTCCNANAATTCNT 628

Qy      266 tggacagtccagatgaaattgacatcaatgtggatgaacttgatacccccgatgaagcag 325
```

```
Db      629 T···NACAACCTGATNACCTGGA···AA.GG.GACNA.ATTAAA···AAA.NAAACTG 685

Qy      326 attcttttgagt 337
            | |||| || |
Db      686 ACTTTTTGGANT 697


RESULT   8
AW217381
LOCUS       AW217381      631 bp    mRNA            EST       06-DEC-1999
DEFINITION  EST296065 tomato flower buds 1-3 mm, Cornell University
            Lycopersicon esculentum cDNA clone cTCAIKi7, mRNA sequence.
ACCESSION   AW217381
VERSION     AW217381.1  GI:6528223
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoe; Lycopersicon.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E.,
            Liang,F., Upton,J., Hansen,T., Craven,M.E., Boatman,D.L., Ann,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Stevens,M.R. and
            Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower buds
  JOURNAL   Unpublished (1999)
COMMENT     On Apr 30, 1999 this sequence version replaced gi:4676447.
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTCAIKi7"
                     /clone_lib="tomato flower buds 1-3 mm, Cornell University"
                     /tissue_type="flower"
                     /dev_stage="0-3mm buds"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Tanksley;  Flower buds and flowers were
                     taken from greenhouse plants (M82 and TA496). They
                     were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      182 a    147 c    [illegible] g    188 t      others
ORIGIN


  Query Match           9.3%;  Score 35;  DB 74;  Length 631;
  Best Local Similarity   55.3%;  Pred. No. 1.9;
  Matches   68;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy      248 tattctctctgatgataacttg[illegible]aatg.actt[illegible] 305
            ||| | |||||| | |  |  |  |  |  |  |  |  |  |  |  |  |
Db      481 TATACAGTCTGATAACAAATCCTCAAAATCAACCAAATCAATA.AAAAATTGTCAA.TT 540

Qy      306 tgataccccgatgaagcagattcttttgagtacactggccatggtaag 365
            |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      541 ATTTGCTTCAGTTCCACTAAAATAGTTTTGATAACTTTAATTTTTTGTTTTAATTTTGC 600

Qy      366 gca 368
            |||
Db      601 TCA 603
```

```
RESULT  9
T11739/c
LOCUS       T11739        243 bp    mRNA            EST       28-NOV-1994
DEFINITION  A1147F Heart Homo sapiens cDNA clone A1147, mRNA sequence.
ACCESSION   T11739
VERSION     T11739.1  GI:596443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 243)
  AUTHORS   Liew,C.C., Hwang,D.M., Fung,Y.W., Laurenssen,C., Cukerman,E.,
            Tsui,S. and Lee,C.Y.
  TITLE     A catalogue of genes in the cardiovascular system as identified by
            expressed sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 10645-10649 (1994)
  MEDLINE   95024171
COMMENT     On Sep 21, 1992 this sequence version replaced gi:279120.
            Other_ESTs: A1147R
            Contact: Liew CC
            Department of Laboratory Medicine and Pathobiology
            University of Toronto
            Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
            Tel: 4169788758
            Fax: 4169785650
            Email: liewcc@utcc.utoronto.ca
            Seq primer: GGTGGCGACGACTCCTGGAGCC.
FEATURES             Location/Qualifiers
     source          1..243
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="A1147"
                     /clone_lib="Heart"
                     /lab_host="E.coli Y1090"
                     /note="Vector: lambda gt11; Site_1: EcoRI; Site_2: EcoRI"
BASE COUNT       54 a     61 c     56 g     72 t
ORIGIN


  Query Match             9.2%;  Score 31.8;  DB 20;  Length 243;
  Best Local Similarity   60.6%;  Pred. No. 1.5;
  Matches   57;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy     281 aaattgacatcaatgtggatgaacttgatacccccgatgaagcagattcttttgagtaca 340
           ||   | ||| || |||||| || | | |   || | | |||||||| | || | | ||
Db     134 AAGCAGCCATTAACGTGGATCAATTCAACATTACCAACGTAGCAGATGCGTTAGTGCACT 45

Qy     341 ctggccatggtaagtcactaagttggcaaggcca 374
           |  |||| || |  || ||  ||| || |
Db      44 TTAACCAAGGCATTCCAACAAAGGCGCAAGGCCA 11


RESULT  10
AQ178668
LOCUS       AQ178668      411 bp    DNA             GSS       17-OCT-1998
DEFINITION  HS_2243_A1_F06_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2243 Col=11 Row=K, genomic survey
            sequence.
ACCESSION   AQ178668
VERSION     AQ178668.1  GI:3576035
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,P., Furlong,T., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG  Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2243  row: K  column: 11
            Class: BAC ends
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1..411
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2243 Col=11 Row=K"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      117 a    103 c     75 g    116 t
ORIGIN


  Query Match             9.0%;  Score 34.2;  DB 99;  Length 411;
  Best Local Similarity   46.4%;  Pred. No. 2.8;
  Matches  111;  Conservative    0;  Mismatches  128;  Indels    9;  Gaps    0;

Qy      29 tctcacccttccgttagaaagcctcctctcatcctcaaaaaactactatcaaaaaatatgtaa 88
           |||   | |  ||   | | |||   ||  || | | | |  |  |  |  |  |  |
Db       3 TCTTCACATAGGTAGTATAACCTGAATTCAGTCTTAAATGGTAAAGTAAATAAGAGTTGT 62

Qy      89 gaatggacatcccctttgaaagaggcctatgagtcccaatactgcaggcataaaaaata 148
           ||||| | |||   ||     || | |   |||||| ||  |  |  |  |  |  |
Db      63 GAATGAAAATGTAATTGACTTATAGTTAAATGATTCTTTTTAACCCTTTTTAAAAAAT 122

Qy     149 aacctcctaatttctctggatcttaatgagactcatcctgaggaatcaaa 208
           | |       |  | | | ||  |  |  ||  |  |  |  |  |
Db     123 CATCCTTCCGGTGGCAGCACCTAAAGTTTCAGGCCCAAAAGACTTTATTTTTTTTTT 182

Qy     209 caaatatcaatctttctctggaccaaagtaaaggatctattctctctgatgatgaaatta 267
           ||    || | | | | |    |  |  ||  || | | |  |  |  |  |  |||
Db     183 ACCGTAATTTTCCTCCACAGAGCTTTGCTATTTTTTCACTTCTGCCATATGGGTTCCTG 241


RESULT  11
Z99377
LOCUS       Z99377        468 bp    mRNA            EST       22-SEP-1997
DEFINITION  HSZ99377 DKFZphamy1 Homo sapiens cDNA clone DKFZphamy1_1b2, mRNA
            sequence.
ACCESSION   Z99377
VERSION     Z99377.1  GI:2415617
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Korn,B., Wiemann,S., Ebert,L. and Poustka,A.
  TITLE     Human ESTs (Korn,B. et al.)
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1144584.
            Contact: B. Korn
            Institution Molekulare Genomanalyse
            Deutsches Krebsforschungszentrum Heidelberg
            Im Neuenheimer Feld 506, D-69120 Heidelberg, FRG.
FEATURES             Location/Qualifiers
     source          1..468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZphamy1_1b2"
                     /clone_lib="DKFZphamy1"
```

```
                     /tissue_type="amygdala"
                     /dev_stage="adult"
BASE COUNT      155 a     84 c    114 g    115 t
ORIGIN


  Query Match              9.0%;  Score 34;  DB 35;  Length 468;
  Best Local Similarity   61.1%;  Pred. No. 3.4;
  Matches   55;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     189 gagaatcaagctcacagccccaaatatcaatctttctctggaccaaagtgaaggatctat 248
           ||||| || || |  || ||| | || |  ||  | ||||| |  |||| || ||| |
Db      85 GAGAAAGAAACTAATGGCTCCAGACATTAGCCTGACACTGGATCCTAGTGATGGCTCTGT 144

Qy     249 tctctctgatgataacttggacagtccaga 278
           | || ||||||       || |||   ||
Db     145 ATTGTCAGATGATTTGGATGAAAGTGGGGA 174


RESULT  12
H11905
LOCUS       H11905         484 bp    mRNA            EST       26-JUN-1995
DEFINITION  ym10c10.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:47661 5', mRNA sequence.
ACCESSION   H11905
VERSION     H11905.1  GI:876725
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:803191.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1806
            High quality sequence stops: 394
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1806   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Homo sapiens"
                     /db_xref="GDB:420202"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:47661"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I, Site_2: Hind III, 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']:
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT      139 a     82 c    103 g    159 t      1 others
ORIGIN


  Query Match              8.8%;  Score 33.2;  DB 22;  Length 484;
  Best Local Similarity   53.0%;  Pred. No. 6;
  Matches   71;  Conservative    1;  Mismatches   63;  Indels    1;  Gaps    0;

Qy     225 ctggaccaaagtgaaggatctattctctctgatgataacttggacagtccaga 285
           |||    ||  | ||||.||||                              .
Db     118 CTGCTTAAAGCTCAAGGATCTTCCTTGAATTTTGTAAATTTGTTCAGTCAAGTTGAAAAG 177

Qy     286 gacatcaatgtggatgaacttgatacccccgatgaaagcagatttttttgagtacaatggc 345
           ' '.    | ||| | :  ||
Db     178 CAAAAACCTATGGTTCACATTCN.!!!!!CNNTGNTNNTTT TTTTNNANANANNNG 237

Qy     346 catggtaagtcact 359
           |'| | ||||: ||
Db     238 TATGATCAGTCTCT 251


RESULT  13
H18541
LOCUS       H18541         469 bp    mRNA            EST       25-JUN-1995
DEFINITION  ym43b03.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:50586 5', mRNA sequence.
ACCESSION   H18541
VERSION     H18541.1  GI:884781
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 469)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 18, 1995 this sequence version replaced gi:[illegible].
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1740
            High quality sequence stops: 347
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1740   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 347.
FEATURES             Location/Qualifiers
     source          1. .469
                     /organism="Homo sapiens"
                     /db_xref="GDB:[illegible]"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:50586"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="[illegible] (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']:
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and Hind III, directly, cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
```

Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 134 a 81 c 97 g 156 t 1 others
ORIGIN

Query Match 8.8%; Score 33.2; DB 23; Length 469;
Best Local Similarity 53.0%; Pred. No. 6;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 226 ctggaccaaagtgaaggatctattctctctgatgataacttggacagtccagatgaaatt 285
Db 118 CTGCTTAAAGCTCAAGGATCTTTCTTGAATTTGTGAAATTTGTTCAGGCAAGGTGAAAAG 177

Qy 286 gacatcaatgtggatgaacttgatacccccgatgaagcagattcttttgagtacactggc 345
Db 178 CAAAAACCTATGGTTCACATTGACTTTTTGTATTGATCATTGTCTTTTGAAGACAGGAAG 237

Qy 346 catggtaagtcact 359
Db 238 TATGATCAGTCTCT 251

RESULT 14
AI417625/c
LOCUS AI417625 463 bp mRNA EST 09-FEB-1999
DEFINITION tq80c11.x1 Soares_NbHMPu_S1 Homo sapiens cDNA clone IMAGE:2115092 3', mRNA sequence.
ACCESSION AI417625
VERSION AI417625.1 GI:4261129
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE 1 (bases 1 to 463)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 7, 1999 this sequence version replaced gi:3120572.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 451.
FEATURES Location/Qualifiers
source 1..463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2115092"
/clone_lib="Soares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ pooled. Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
BASE COUNT 154 a 87 c 76 g 145 t 1 others
ORIGIN

Query Match 8.8%; Score 33.2; DB 45; Length 463;
Best Local Similarity 53.0%; Pred. No. 5.9;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 226 ctggaccaaagtgaaggatctattctctctgatgataacttggacagtccagatgaaatt 285
Db 368 CTGCTTAAAGCTCAAGGATCTTTCTTGAATTTGTGAAATTTGTTCAGGCAAGGTGAAAAG 309

Qy 286 gacatcaatgtggatgaacttgatacccccgatgaagcagattcttttgagtacactggc 345
Db 308 CAAAAACCTATGGTTCACATTGACTTTTTGTATTGATCATTGTCTTTTGAAGACAGGAAG 249

Qy 346 catggtaagtcact 359
Db 248 TATGATCAGTCTCT 235

RESULT 15
AW104968/c
LOCUS AW104968 402 bp mRNA EST 2-NOV-1999
DEFINITION xd73d06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2603243 3', mRNA sequence.
ACCESSION AW104968
VERSION AW104968.1 GI:6075703
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5866157.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 379.
FEATURES Location/Qualifiers
source 1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2603243"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ pooled. Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B cell NCI_CGAP_GC6), were mixed, and ss circles made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 247490-252787, 692632-697284, 726488-728711, and 729096-731999, subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 136 a 72 c 59 g 135 t
ORIGIN

Query Match 8.8%; Score 33.2; DB 45; Length 469;
Best Local Similarity 53.0%; Pred. No. 5.6;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 226 ctggaccaaagtgaaggatctattctctctgatgataacttggacagtccagatgaaatt 285
Db 373 CTGCTTAAAGCTCAAGGATCTTTCTTGAATTTGTGAAATTTGTTCAGGCAAGGTGAAAAG 314

Qy 286 gacatcaatgtggatgaacttgatacccccgatgaagcagattcttttgagtacactggc 345

```
Db    313 CAAAAACCTATGGTTCACATTGACTTTTTGTATTGATCATTGTCTTTTGAAGACAGGAAG 254

Qy    346 catggtaagtcact 359
          ||| | |||| ||
Db    253 TATGATCAGTCTCT 240


Search completed: February  7, 2000, 19:33:33
Job time: 6723 sec
```

```
OM of: US-09-174-937-5 to: EST:*  out_format : pfs

Date: Feb 10, 2000  1:33 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09174937/runat_10022000_095107_13273/app_query.fasta.1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -FRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-174-937-5
Query length: 322
Database: EST:*
Database sequences: 4538634
Database length: 1887831982
Search time (sec): 1131.150000



score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation ..
gb_est1:T31973        +  508.00 1099.16  1.9e-52   423  ! T31973 EST41767 Human Uterus Ho
gb_est27:AI437189     -  480.00 1035.52  6.5e-49   439  ! AI437189 fb38g06.y1 Zebrafish W
gb_est1:T05645        +  466.00 1008.28  2.1e-47   369  ! T05645 EST03534 Fetal brain, St
gb_est15:AA504063     -  461.00 1000.71  5.7e-47   275  ! AA504063 nh40g09.s1 NCI_CGAP_Pr
gb_est10:AA137873     -  459.50  991.61  1.8e-46   449  ! AA137873 mq75d03.r1 Stratagene
gb_est14:AA450831     -  447.50  967.53  4.0e-45   368  ! AA450831 vg55e03.r1 Boddington
gb_est1:Z32816        -  430.00  930.92  4.1e-43   312  ! Z32816 HHEA02P Atrium cDNA libr
gb_est9:AA117114      -  425.50  918.60  2.1e-42   392  ! AA117114 zn68h02.r1 Stratagene
gb_est32:AI722140     -  399.00  865.00  6.6e-38   604  ! AI722140 fd29g03.y1 Sugano Kawa
gb_est21:AA914577     -  397.00  830.25  1.3e-37   516  ! AA914577 vz02e07.r1 Soares mous
gb_est29:AI620423     -  372.00  797.91  1.1e-35   486  ! AI620423 tu47b05.x1 NCI_CGAP_Pr
gb_est12:AA296475     -  366.00  793.23  2.0e-35   236  ! AA296475 EST10971 Umbilical Vei
gb_est9:AA067188      -  358.00  767.93  5.2e-34   452  ! AA067188 mm38b12.r1 Stratagene
gb_est9:AA068579      -  358.00  768.17  5.1e-34   443  ! AA068579 mm45e06.r1 Stratagene
gb_est1:T31983        +  354.50  765.84  6.8e-34   291  ! T31983 EST41923 Human Uterus Ho
gb_est10:AA129390     -  337.50  724.39  1.4e-31   394  ! AA129390 zr84e05.r1 Stratagene
gb_est5:D61567        +  336.00  724.23  1.4e-31   302  ! D61567 HUM418G12B Clontech huma
gb_est2:T68254        -  306.50  654.45  1.1e-27   403  ! T68254 yo40g01.r1 Stratagene li
gb_est2:T94657        -  304.00  649.26  2.1e-27   441  ! T94657 ye37e08.r1 Stratagene li
gb_est24:AI180114     -  289.00  614.13  1.9e-25   526  ! AI180114 EST223851 Normalized s
gb_gss1:FR0014613     -  278.00  589.18  4.7e-24   559  ! AL005822 F.rubripes GSS sequenc
gb_est37:AI976981     -  256.00  541.59  2.1e-21   519  ! AI976981 EST271498 Schistosoma
gb_est12:AA280204     -  250.00  529.62  4.8e-21   467  ! AA280204 zt05c05 r1 NCI_CGAP_Co
gb_est22:AI006991     -  241.00  512.97  9.7e-20   257  ! AI006991 ua78a04.r1 Soares 2NbM
gb_est21:AA980800     -  234.00  494.10  6.7e-19   478  ! AA980800 ua45g06.r1 Soares mous
gb_est39:AW106241     -  230.00  491.17  4.0e-19   677  ! AW106241 um25a08.y1 Sugano mous
gb_est14:AA396890     -  206.00  432.69  2.5e-15   467  ! AA396890 mr38g06.r1 Life Tech m
gb_est1:T32180        -  203.00  436.97  1.1e-15   186  ! T32180 EST44614 Human Brain Hum
gb_est16:Z99377       +  199.50  418.35  1.5e-14   468  ! Z99377 HSZ99377 DKFZphamy1 Homo
gb_est21:C91038       +  199.50  414.61  2.5e-14   612  ! C91038 C91038 Dictyostelium dis
gb_est32:AI727670     -  193.00  401.87  1.3e-13   562  ! AI727670 BNLGH18551 Six-day Cut
gb_est10:AA129317     -  189.00  393.29  3.0e-13   551  ! AA129317 zr84e05.s1 Stratagene
gb_est38:AW052742     +  188.00  392.74  4.1e-13   577  ! AW052742 614031B10.y1 614 - rad
gb_gss1:FR0009132     -  177.50  367.51  1.1e-11   572  ! AL000428 F.rubripes GSS sequenc
gb_est20:AA866746     +  174.50  360.96  2.4e-11   569  ! AA866746 ud07e04.r1 Soares mous
gb_est44:AW201496     -  174.00  359.55  3.9e-11   594  ! AW201496 sf02g01.y1 Sugano mous
gb_est33:AI781944     -  172.00  354.48  5.6e-11   618  ! AI781944 EST262223 tomato susce
gb_est35:AI916929     -  170.50  351.52  8.2e-11   600  ! AI916929 wj35d01.x1 NCI_CGAP_Ki
gb_est42:AV293206     -  169.00  348.52  1.2e-10   212  ! AV293206 AV293206 RIKEN full-le
gb_est9:AA120357      -  165.00  342.48  2.3e-10   425  ! AA120357 mp89f04.r1 Soares 2NbM
gb_est32:AI731622     +  164.00  337.65  4.9e-10   578  ! AI731622 BNLGH10281 Six-day Cu
gb_est16:AA578186     -  159.00  334.87  6.9e-10   288  ! AA578186 nl29e10.s1 NCI_CGAP_HS
gb_est28:AI517653     -  159.00  334.71  1.0e-09   548  ! AI517653 wm26d96.5prime GI
gb_gss11:AQ018285     -  159.00  334.71  1.2e-09   555  ! AQ018285 RPCI11-108A18.TJ
gb_gss11:AQ018293     -  159.00  334.19  1.4e-09   541  ! AQ018293 RPCI11-108A20.TJ

seq_name: gb_est1:T31973

seq_documentation_block:
LOCUS       T31973          423 bp    mRNA            EST       6-SEP-1995
DEFINITION  EST41767 Human Uterus Homo sapiens cDNA 5' end similar to similar
            to rhoGAP protein, mRNA sequence.
ACCESSION   T31973
VERSION     T31973.1  GI:614071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chin,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.F., Li,J., Mahairas,S.M.,
            Merrick,J.M., Moreno-Palanques,R.E., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.P., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,J., Coppola,M.A.,
            Coleman,T.A., Collins,E.J., Limke,D., Ford,P., Ferrio,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,C.-S., Greene,J.M.,
            Gruber,J., Hudson,F., Kim,A., Kozak,C.L., Kuhsen,G., Li,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Weiland,T., Wing,J., Xu,C.,
            Yu,J.-L., Zuberi,H.M., Allen,P.G., Taussig,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expressed Patterns
            Based Upon 83 Million Nucleotides of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026180
COMMENT     Other_ESTs: IMAGE:4305
            Contact: Venter, JC
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1..423
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Uterus"
                     /note="Organ: uterus"
BASE COUNT      126 a      71 c      90 g     131 t      3 others
ORIGIN


alignment_scores:
            Quality:  508.00              Length:    145
              Ratio:  3.938                 Gaps:      4
 Percent Similarity:  86.906   Percent Identity:  84.058

alignment_block:
 US-09-174-937-5 x T31973  ..

 Align Seq 1/1  to: T31973  from: 1  to: 423

     159 IleGlnProTyrArgArgValIle...GlyAspAsnGlyTyrTyr 177
         ||||:|||:|||||:::::|||
```

```
      5 ATTGAACCCTATAAAAAAGTTATCAGCCATGGG.........GGATATTA 45

    175 rGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProA 192
        |   |||:|||||||||||||:::|||||:||  |:::|||||||:::|||
     46 TNGGGATGGATTAAATGCCATTGTTGTGTTTGCTGCTGTTTCATGCCTG 95

    192 spSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyr 208
        ::|||||:::   :::|||:::|||:::||:::|||||||   |||
     96 AAAGTAGTCAGCCTAACTATAGATACCTGATGGACAATCTTTTTAAATAT 145

    209 ValIleSerThrLeuGlyLeuMetValAlaGlyAspTyrMetIleValTy 225
        |||||:::'|||'||||||||||::|||||'||  :::|||||||||||||
    146 GTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAACTACATGATAGTTTA 195

    225 rLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetL 242
        |'|'|'|'|||||||   '||||:::| ||  :::|||||||||:::::
    196 TTTAAATGGTGCAACAACTGGAAGAAAAATGCCCAGTCTGGGATGGCTCA 245

    242 ysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSe 258
        ::|||||||||||||:::|||||||||||||||||'|| ||||||||||||||
    246 GGAAATGTTATCAGCAAATTGATAGAAGGTTACTGGAAAAATCTAAAATC 295

    259 rPheIleIleValHisProSerTrpPheIleArgThrIleLeuAlaValT 275
        |:::|||||||||||||||||||   |||| ||||...||||||||.
    296 CCTAATCATTGTACATCCTTCTTGGTTTATCAGAACACTTCTGGCTGTTA 345

    276 hrArg.ProPheIleSerSerLysPheSerSerLysIleLysTyrValAs 291
        ::|||  .'|'|||||||||:::|||:::   '||||   -||   ::
    346 NAAGGCCTTTTATTAGCTGGAATTTCAGCAAAAATTTGGTAC   GG 392

    291 nSerLeuSerGluLeuSerGlyLeuIlePro 301
        :   ::::::'||   ||'   '||:::'||
    393 GTTTATTTNGGAGACTAGCAGACTTGTCCCC 423

seq_name. gb_est27:AI437189

seq_documentation_block:
LOCUS       AI437189      499 bp    mRNA            EST       09-MAR-1999
DEFINITION  fb38g06.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
            TR:Q12992 Q12992 NIP2. , mRNA sequence.
ACCESSION   AI437189
VERSION     AI437189.1  GI:4288571
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinidae; Cyprininae; Danio.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,R., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,F.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138723.
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham

            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1..499
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']:
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab: ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to statistically cluster cDNAs and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      163 a    110 c    117 g    109 t
ORIGIN


alignment_scores:
            Quality:  480.00              Length:    119
              Ratio:  4.286                 Gaps:      0
 Percent Similarity:  94.118   Percent Identity: 68.908

alignment_block:
 US-09-174-937-5 x AI437189   ..

 Align seg 1/1  to: AI437189  from  1  to: 499

    197 AspTyrHisTyrValMetGluAsnLeuPheLeuTyrValIleSerThrLe 213
        ||||||||||||||:::|||||||||| |:|||||'|||||:::'|:::::|
      6 GATTACCATTACCTGATGGAAAATCTCTTTCTGTATGTTGTAAGCAGTTT 55

    214 uGluLeuMetValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaT 230
        ||||  :::|||||||||||||||| ::|||:::|||:::|||||||:::|
     56 GGAGATGCTGGTGGCGGAGGATTACTGATCATCTACATGAATGGAGCAA 105

    231 hrProArgArgArgMetProGlyLeuGlyTrpMetLysLysCysTyrGln 246
        |||||||      ||||||||||:: ::::|:::||||||||:|:|,,
    106 CGCCGCGGAGTAAAATGCCCGGCATTAGTTGGTCAAGAAGTGCTACCAG 155

    247 MetIleAspArgArgLeuArgLysAsnLeuLysSerPheIleIleValHi 263
        |||||||||||||||||||||||| ||||||||||::: :: |::::
    156 ATGATCGACAGAAGACTGAGAAAAAACTTAAAATCTCTGATCATCATTTA 205

    263 sProSerTrpPheIleArgThrIleLeuAlaValThrArgProPheIlos 28
        |||||||||||||||||||||||      |||     ...||||||||||
    206 TCCCTCCTGGTTTATTCGGACGGTCATCGCCATTCCAAACCGTTCATCA 255

    280 erSerLysPheSerSerLysIleLysTyrValAsnSerLeuSerGluLeu 29
        ||   ||||||   :::|||||||:::'|'|.|:::|||||'::::.|'
    256 GCGTGAAGTTCATGAATAAGATCTGTTACGTGCACAGTTTGTAACAGCTT 305

    297 SerGlyLeuIleProMetAspCysIleHisIleProGluSerIleIleAs 313
                ||||||||   ||||:|||||||||   ||:::::
    306 GAAAAGTTCGTTCCCATGGATCATATTCATATCCCAGAATCCATCTTACA 355

    313 nIleAsp 315
        :  :::
    356 ATATGAA 362
```

```
seq_name: gb_est1:T05645

seq_documentation_block:
LOCUS       T05645        368 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST03534 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDD71, mRNA sequence.
ACCESSION   T05645
VERSION     T05645.1  GI:316795
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3019694423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):82344"
                     /db_xref="taxon:9606"
                     /clone="HFBDD71"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                     vector; 1.0kb average insert size."
BASE COUNT       88 a    106 c     90 g     80 t      4 others
ORIGIN


alignment_scores:
            Quality:  466.00             Length:    119
              Ratio:  4.198                Gaps:      0
 Percent Similarity:  93.277   Percent Identity:  78.950

alignment_block:
 US-09-174-937-5 x T05645   ..

 Align seq 1/1  to: T05645  from: 1  to: 368

     197 AspTyrHisTyrValMetGluAsnLeuPheLeuTyrValIleSerThrLe 213
         ||||||||||||:::|||||||||||||||||||||||||||||||:::||
       2 GACTACCACTACATCATGGAGAACCTCTTCCTGTACGTNATCAGCACGCTT 51

     213 uGluLeuMetValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaT 230
         |||||||:::|||||||||||||||||||| |  ||||||   ||||||||
      52 AGAGCTCCTGGTGGCTGAGGACTACATGATCGTNACCTGANCGGTGCCA 101

     230 hrProArgArgArgMetProGlyLeuGlyTrpMetLysLysCysTyrGln 246
         |||| |||| |||||| | || ||||| |  ::::|||||||||||||||
     102 CGCCCCGGCGGAGGATGCCTGGAATCGGCTGGTGAAGAAGTGCTACCAG 151

     247 MetIleAspArgArgLeuArgLysAsnLeuLysSerPheIleIleValHi 263
         |||||||||| |  ||| |||||||||||| |  ||:::|||||||||||
     152 ATGATCGACCGGAG.TTGCGGAAAAACCTGAAGTCCTTGATCATCGTCCA 200

     263 sProSerTrpPheIleArgThrIleLeuAlaValThrArgProPheIleS 280
         |||||||||||||||||||||||:::|||||||:::::|||||||||||||
     201 CCCCTCGTGGTTCATTCGGATTGTGGTGGCCA.TTCTCGCCCTTTCATCA 250

     280 erSerLysPheSerSerLysIleLysTyrValAsnSerLeuSerGluLeu 296
         ||   ||||||   :::|||||||:::|||.|:::|||||||:::::|||
```

```
     251 GCGTCAAGTTCATCAACAAGATCCATTATGTGTACATCTTTGTAATATTTN 300

     297 SerGlyLeuIleProMetAspCysIleHisIleTrpTyrPheLysAsN 343
         ::: :::::  ||| ::::::  ::: ::::::::  ||||||||||| |
     301 GAGCAACTCATCCCTATGGAACATTTCATATTTCATATTTGTTTTTTA 347

     313 nIleAsp 315
         :  :::
     351 ATACGAG 357

seq_name: gb_est15:AA504063

seq_documentation_block:
LOCUS       AA504063      275 bp    mRNA            EST       [illegible]-APR-1997
DEFINITION  nh40c09.s1 NCI_CGAP_Pr2 [illegible] Homo sapiens cDNA [illegible] IMAGE:[illegible], mRNA
            sequence.
ACCESSION   AA504063
VERSION     AA504063.1  GI:2239003
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   NCI-CGAP; http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1999 this sequence version replaced gi:[illegible].
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Bradley P. Munaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David R. Krizman, Ph.D.
            cDNA Library Arrayed by: Bier Lehman, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Cons [illegible]/UNIGENE at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 404  Std Error: 0.00
            Seq primer: -40m13 fwd, RT from Amersham
            High quality sequence stop: 214.
FEATURES             Location/Qualifiers
     source          1..275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:[illegible]"
                     /clone_lib:"NCI_CGAP_Pr5"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA from [illegible] prostatic
                     epithelial cells; cDNA made by oligo dT primed;
                     Non-directionally cloned; Size-selected [illegible] agarose gel;
                     average insert size [illegible] bp."
BASE COUNT       [illegible] a     [illegible] c     [illegible] g     [illegible] t
ORIGIN


alignment_scores:
            Quality:  461.00             Length:
              Ratio:  5.122                Gaps:
 Percent Similarity:  98.901   Percent Identity:  [illegible]

alignment_block:
 US-09-174-937-5 x AA504063   ..

 Align seq 1/1  to: AA504063  from: 1  to: 275

      15 SerArgProAsnGlyLeuLeuSerIleAspValGlyMetAspIleProPh 41
         :::  ||||.|||||:.| |
```

```
      3 TCGACACCAAAATGGACTACTGTCAGAGGATGTACGAATGGACATCCCCTT 52

     31 eGluGluGlyValLeuSerProSerAlaAlaAspMetArgProGluProP 48
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     53 TGAAGAGGGCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTC 102

     48 roAsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThr 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    103 CIAATTCTCTGGATCTTAATGACACTCATCCTCGGAGAATCAAGCTCACA 152

     65 AlaProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSe 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    153 GCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTC 202

     81 rAspAspAsnLeuAspSerProAspGluIleAspIleAsnValAspGluL 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    203 TGATGATAACTTGGACAGTCCAGATCAAATTGACATCAATGTGGATGAAC 252

     98 euAspThrProAspGluAlaAsp 105
        ||||::||||||||||||||||
    253 TTGATAGCCCCGATGAAGCAGAT 275

seq_name: gb_est10:AA137873

seq_documentation_block:
LOCUS       AA137873      449 bp    mRNA            EST       09-FEB-1997
DEFINITION  mq75d03.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:584540 5' similar to TR:G558844 G558844 NIP2  ;, mRNA
            sequence.
ACCESSION   AA137873
VERSION     AA137873.1  GI:1700244
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394228.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1900
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information
            MGI:359197
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:584540"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SURF (streptomycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: -28 adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      147 a     76 c    124 g    102 t

ORIGIN


alignment_scores:
           Quality:  459.50              Length:     128
             Ratio:  4.066                 Gaps:       2
 Percent Similarity:  88.281   Percent Identity: 86.406

alignment_block:
 US-09-174-937-5 x AA137873   ..

 Align seg 1/1  to: AA137873  from: 1  to: 449

     126 SerIleProGluTyrThrAlaGluGluGluArgGluAspAsnArgLeuTr 142
         ||||||  ||||||||||||  |||| ||||  ||||||||||
      75 TCGATTACTGAGTACACAGCCACAGAAGAAAAGGGCGATGGACGCCGCTG 124

     142 pMetThrValValIleGlyGluGlnGluGlnArgIleAspMetLysValI 191
         |          ||||||||| ||||||||||||||||  |  |||
     125 GCCAATGTTCAGGATTGGAGAGCAGCAGCACAGGGTGGACATGAAGGCGA 174

     159 leGluProTyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTyr 207
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 TCGAGCCCTACAAAAAAGTTATTAGCCATGGA.........GGATATTAT 217

     176 GlyAspGlyLeuAsnAlaIleIleVal PheAlaAlaCysPheLeuProA 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 GGGGACGGCTTAAATGCCATTGTTGTATTTGCTGCTGCTTTCATGCCTG 267

     192 spSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyr 208
         ::||||::::::   :::|||:::||::::| |:::|||||||:||   ||
     268 AGAGTGGTCAGCCTAACTATAGATATTTGATGGACAATCTCTTTAAATAT 317

     209 ValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValTy 225
         ||||||...|||||||||||||.:|||||||||||||||||||||||||
     318 GTTATTGGCACTTTAGAGCTGTTGGTAGCAGAACTACATGATCATTTA 365

     225 rLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetL 242
         ||||||||||||||||   |||||:::|||||||:::|||||||||:::
     366 CTTAAATGGTGCGACAACTCGAAGAAAAATCCCAGTTTGGGATGGTTCA 415

     242 ysLysCysTyrGlnMetIleAspArgArgLeu 252
         :::::|||||||||:::|||||||| ||||||
     416 GGAGATGCTACCAGCAAATTGATAGACGATTA 447

seq_name: gb_est14:AA450831

seq_documentation_block:
LOCUS       AA450831      368 bp    mRNA            EST       13-JUN-1997
DEFINITION  vg55e03.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865276 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA450831
VERSION     AA450831.1  GI:2164501
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1947[illegible].
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO [illegible]
            Tel: 314 286 1800
```

```
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509364
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Mus musculus"
                     /strain="C57BL6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:865276"
                     /clone_lib="Beddington mouse embryonic region"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT  Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings, excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                     Referenced in Development 121, 2479-2489 (1995)"
BASE COUNT      110 a     66 c     98 g     94 t
ORIGIN


alignment_scores:
            Quality:  447.50              Length:    126
              Ratio:  3.996                 Gaps:      1
 Percent Similarity:  88.889   Percent Identity: 86.667

alignment_block:
 US-09-174-937-5 x AA450831   ..

 Align seq 1/1  to: AA450831  from: 1  to: 368

     126 SerIleProGluTyrThrAlaGluGluGluArgGluAspAsnArgLeuTr 142
         ||||||  ||||||||||||  ||||||:::  |||:::|||  ||
       1 TCGATTACTGAGTACACAGCCACAGAAGAAAAAGGGCGATGGACGCCGCTG 50

     142 pMetThrValValIleGlyGluGlnGluGlnArgIleAspMetLysValI 159
         |           |||||||||||||:::::|||:::|||||||||:::|
      51 GCC.ATGTTCAGGAATTGGAGAGCAGGACCACAGGGTGGACATGAAGGCGA 99

     159 leGluProTyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTyr 175
         ||||||||||:::::|||||||||||||||||         |||||||||
     100 TCGAGCCCTACAAAAAAGTTATTAGCCATGGA.........GGATATTAT 140

     176 GlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAs 192
         |||||||||||||||||||||||:::|||||||| ||||||||:::|||::
     141 GGGGACGGCTTAAATGCCATTGTTGTATTTGCTGTCTGTTTCATGCCTGA 190

     192 pSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyrV 209
         ::|||:::::  :||||  ||||:::|  ||:::|||||||||  ||||
     191 GAGTGGTCAGCCTAACTATAGATATTTGATGCAACAATCTCTTTAAATATG 240

     209 alIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValTyr 225
         ||||:::|||||||||||||||:::|||||||:::|||||||||:::|||
     241 TTATTGGCACTTTAGAGCTGTTGGTAGCAGACAACTACATGATCATTTAC 290

     226 LeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLy 242
         ||||||||||||||  ||||||  ||||||  ||||||||||:::::
     291 TTAAATGGTGCGACAACTCGAAGAAAAAATGCCTAGTTTGGGATGGCTCAG 340

     242 sLysCysTyrGlnMetIleAspArgArg 251
         :::|||||||||||:::|||||||||||
     341 GAGATGCTACCAGCAAATTGATAGACGA 368

seq_name: gb_est1:Z32816
```

```
seq_documentation_block:
LOCUS        Z32816        312 bp    mRNA            EST       8-MAY-1994
DEFINITION   HHEA02P Atrium cDNA library, Human heart Homo sapiens cDNA clone
             HEA02P, mRNA sequence.
ACCESSION    Z32816
VERSION      Z32816.1  GI:483499
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 312)
  AUTHORS    Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
             Devignes,M.D., Duprat,S., Houlgatte,R., Jimeno,M.N., Lamy,B.,
             Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Poiliot,Y.,
             Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE      IMAGE: molecular integration of the analysis of the human genome
             and its expression
  JOURNAL    C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE    95277534
COMMENT      On Sep 21, 1992 this sequence version replaced gi:278108.
             Contact: Genzentrum Muenchen
             Laboratorium fuer molekulare Biologie
             Am Klopferspitz 19a,8033 Martinsried,Germany
             Email: obermaier@vms.biochem.mpg.de
             single read.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEA02P"
                     /clone_lib="Atrium cDNA library Human heart"
BASE COUNT       97 a     55 c     59 g    101 t
ORIGIN


alignment_scores:
            Quality:  430.00              Length:    1 9
              Ratio:  4.343                 Gaps:      .
 Percent Similarity:  96.117   Percent Identity: 7 .7 %

alignment_block:
 US-09-174-937-5 x Z32816   ..

 Align seq 1/1  to: Z32816  from: 1  to: 312

     182 IleIleValPheAlaAlaCysPheLeuProAspSerSerArgAlaAspTy 198
         ||:::|||||||||:::|||||  :::  :::  :::  :::  :::
       1 ATTTTGTGTTTGCTGTCTGTTTCATGCCTGAAAGTAGTCAGCCTAATTA 50

     198 rHisTyrValMetGluAsnLeuPheLeuTyrValIleSerThrLeuGlu 215
         |...  |...|:::|||||||  '  |  '  |||  :::
      51 TAGATACCTGATGGAGAATTTTTTAAATATGTTATTGGCACTTTGGAGC 10.

     215 euMetValAlaGluAspTyrMetIleValT,rLeuAsnGlyAlaThrPro 2.1
         | :::|||||||||:::|||||  '  |  '  |  '  |  :::  :::
     101 TACTAGTAGCAGAAAACTACATGATAGTTTATTTAAATGGTGCAACAACT 15.

     232 ArgArgArgMetProGlyLeuGlyTrpMetLysLysLysLysMetIl 24-
         |||'|:::|'||:|:|||  :::::  :::  :::  :::  :::
     151 CGAAGAAAAAATGCCCAGTCTGGGATGGGTCAGTAATGTTATCAATCAAAAT 2..

     248 eAspArgArgLeuArgLysAsnLeuLysSerPheIleIleVal.HisProS 26.
         ||||||||||||||||:::|||  :::  :  :  :  :  :  :  :
     201 TGATAGAAGGTTACGGGAAAAICTAAAATCCCTAATCATTGTACATCCTT 25.

     265 erTrpPheIleArgThrIleLeuAlaValThrArgProPheIleSerSer 28.
         '|||'||||||||||| :::|  :  :  :  :  :  :  :  :  :
     251 CTTGGTTTATCAGAACACTTCTGGGCTGTTACAAGACCATTTATTAGCTTG 30.

     282 LysPheSer 284
         |||  |'||
     301 AAATTCAGC 309
```

```
seq_name: gb_est9.AA113114

seq_documentation_block:
LOCUS       AA113114      382 bp    mRNA            EST       01-AUG-1997
DEFINITION  zn68h02.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            IMAGE:563379 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA113114
VERSION     AA113114.1  GI:1664694
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On May 5, 1995 this sequence version replaced gi:797955.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1900
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2034  Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1..382
                     /organism="Homo sapiens"
                     /db_xref="GDB:459319"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:563379"
                     /clone_lib="Stratagene HeLa cell s3 937216"
                     /sex="female"
                     /dev_stage="HeLa S3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally; Primer: Oligo dT, HeLa S3
                     epithelioid carcinoma cells grown to semi-confluency
                     without induction; Average insert size: 1.5 kb; Uni-ZAP XR
                     Vector; 5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      131 a     59 c     93 g     99 t
ORIGIN


alignment_scores:
            Quality:  425.50            Length:    120
              Ratio:  3.977               Gaps:      3
 Percent Similarity:  89.167   Percent Identity: 68.333

alignment_block:
 US-09-174-937-5 x AA113114   ..

 Align seq 1/1  **  AA113114  from  1  to  382

     126 SerIleProGluTyrThrAlaGluGluGluArgGluAspAsnArgLeuTr 142
         ||||||  ||||||||||||  ||||||::: |||||::|||  ||
      35 TCAATTACTGAATACACAGCAGCAGAGGAAAAAGAAGATGGACGACGCTG 84

     142 gMetThrValVal GluGlyGlyGlyGlyArgIleAspMetLysVal 158
         |  :::|||   ||||||||||||||:     |  ||||||||  :
```

```
      85 GCGTAGTGTTCAGGATTGGAGAACAGGACCACAGGGTTGATATGAAGTTA 134

     159 IleGluProTyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTy 208
         ||||||||||||:::|||||||||||||          ||||
     135 ATTGAACCCTATAAAAAAGTTATTAGCCACGGC          TGATATTA 178

     175 rGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProA 224
         ||||||||||||||||||||||:::|||||||:::|||||:::| ||
     179 TGGGGATGGATTAAATGCCATTGTTGTGTTTGCTGTCTGTTTCATGCCTG 228

     192 spSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyr 241
         ::|||||...  ..|||:::|||:::|||:::|||||||||  ||
     229 AAAGTAGTCACCCTAACTATAGATACCTGATGGACAATCTTTTTAAATAT 278

     209 ValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValTy 258
         ||||||:::|||||||||||||:::|||||||||:::|||||||||||
     279 GTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAACTACATGATACTTTA 328

     225 rLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGly.TrpMet 274
         ||||||||||||||||  ||||||:::|||||||:::|||||| ||||:
     326 TTTAAATGGTGCAACAACTCGAAGAAAAATGCCCAGTCTGGGAATGGCTG 375

     242 LysLys 243
         ::||||
     376 AGGAAA 381

seq_name: gb_est32.AI722140

seq_documentation_block:
LOCUS       AI722140      604 bp    mRNA            EST       [illegible] MAY-1999
DEFINITION  fd19g03 y1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
            2352052 5' similar to TR:O15072 O15072 KIAA0267 [illegible] mRNA sequence.
ACCESSION   AI722140
VERSION     AI722140.1  GI:5040469
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S.L., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Stapleton,M., Underwood,K., Theising,B.,
            Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:[illegible].
            Other_ESTs: fd19g03.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2352052"
                     /clone_lib="Sugano Kawakami zebrafish DNA"
                     /sex="mixed (one male and one female) including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: BraIII (CATTTTG);
                     Site_2: BraII (CACCATGC), 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTT];
```

double-stranded cDNA was ligated to a DraIII adaptor [IGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT 184 a 145 c 153 g 121 t 1 others
ORIGIN

alignment_scores:
Quality: 390.00 Length: 171
Ratio: 3.095 Gaps: 8
Percent Similarity: 73.684 Percent Identity: 53.216

alignment_block:
US-09-174-937-5 x AI722140 ..

Align seq 1/1 to: AI722140 from: 1 to: 604

```
      22 SerGluAspValGlyMetAspIleProPheGluGluGlyValLeuSerPr 38
         |||   :::|||||||||||||||      :::   ||||...   ...|||
     126 TCTGGCAAATGTGGGGATGGACCAGACTTAT...GAAAGCAAGATGAGTTT 172

      38 oSerAlaAlaAspMetArgProGluProProAsnSerLeuAspLeu.... 53
          :::::::::::   ||||||   |||||||::||||||   |||
     173 TGAAGGGGCAGAGAGTAGACCAGCTGCACCCAGCTCACTGCCTCTACAGG 222

      54 .....AsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIle 68
           :::   ...     ...|||   ||| |||...||||||:::|||
     223 GCCAAGGGGGTGTGTCTTTCAGAGGAAAAAACTGTCCGTTCCCCGTATT 272

      69 AsnLeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLe 85
         :::|||||||||||||||||||||                ||||||   ::
     273 AGTCTTTCATTGGACCAAAGTGAG.              .... GATGACCTGTT 307

      86 uAspSerProAspGluIleAspIleAsnValAspGluLeuAspThrProA 102
         :|||:..|||:|||.....:|||||||||||||||:|.:::|||||||||||||
     308 TGACACGCCAGATGACCTCCACATCAATGTGCAGGATCTGGACACACCTG 357

     103 spGluAlaAspSerPheGluTyrThrGlyHis..............Asp 113
         ||||||||||||          '||||||   |'|                :::
     358 ATGAAGCAGAT... .     TATACAGACCATGAAATAGACTGGGAAGAA 399

     114 ProThrAlaAsn   LysAspSerGlyGlnGluSer...GluSerIlePr 128
         |||   |||:::   :::   |||   ...:|||:::   |||   |||||
     399 CCTCAAGCATCCCAGCGAGTTTCTGTAAAAGAAACAGTTGAGCCCATTCC 449

     129 oGluTyrThrAlaGluGluGluArgGluAspAsnArgLeuTrpMetThrV 145
         |   |||:::'||||||||| ||||:::|| ::::::|||||||   ||||
     449 CACATACAGTGCTGAGGAGGAAGGGCAGGATTCCAAACTCTGGAGAACGG 498

     145 alValIleGlyGluGlnGluGlnArgIleAspMetLysValIleGluPro 161
         |||:::|||||||||||||||||:::|||||||:::| ||||   |||||||||
     499 TCATCATTGGAGAGCAAGAGCAGCGGATCAACATGAAGAGCATTGAACCC 548

     162 TyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAspGl 178
         |||:::::::::::::::::::::                ||||||||||:::::|||
     549 TACCAAAAAGTCATAATCCCATGGA.  .       TGCTACTATAGCAATGG 589

     178 yLeuAsnAlaIle 192
         |      ||||||
     590 TGCAAATGCCATC 602

seq_name: gb_est21:AA914577

seq_documentation_block:
LOCUS AA914577 516 bp mRNA EST 14-APR-1998
DEFINITION vz02e07.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
IMAGE:1314564 5' similar to TR:Q14842 Q14842 NIP1. mRNA
sequence.
ACCESSION AA914577
VERSION AA914577.1 GI:3053969
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 516)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,C., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,M., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1998 this sequence version replaced gi:1314564.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:684860
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 519.
FEATURES Location/Qualifiers
source 1..516
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:...9."
/clone="IMAGE:1314564"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3-Pac
(Pharmacia, with a modified polylinker site); Not I:
Site_2; Eco RI: 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
IGTTACCAATC]GAAG]GGAGCGGCGCAAA]TTTTTTTTT]TTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3-Pac vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 148 a 110 c 118 g 141 t
ORIGIN

alignment_scores:
Quality: 387.00 Length: 165
Ratio: 4.074 Gaps: 1
Percent Similarity: 90.476 Percent Identity: 68.571

alignment_block:
US-09-174-937-5 x AA914577 ..

Align seq 1/1 to: AA914577 from: 1 to: 516

```
     220 AspTyrMetIleValTyrLeuAs.GlyAlaTh:rGlnArgArgMetPr 24
         ::: | |'|'| :::'|'|||'' ' ' ' :::
       2 AACTACATGATCATTTACTTAAA:GG:GGCAACTGCAAGAAAAATGG 51

     236 GGlyLeuGlyTrpMetLysLysGlyGlyGluMetIleAspArgArgLeuA 25
         |:::|||||||||:::::::::' ' ' :::' |
      52 CAGTTTGGGATGGCTCAGGAGATTTTCGCAGCAAATTGATAGAGATTAA 1
```
```

```
     253 rgLysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArg 269
         ||||||| |||||||||||:::|||||||| || |||||||||||||||
     102 GGAAAAACCTTAAGTCTCTAATCATTGTGCACCCTCTTGGTTTATCCGA 151

     270 ThrIleLeuAlaValThrArgProPheIleSerSerLysPheSerSerLy 286
         ||:::|||||||||||||||||||||||| || ||||||||||:::||
     152 ACACTTCTGGCTGTTACAGGACCATTTATCAGTCCAAAGTTCAGTCAGAA 201

     286 sIleLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMetA 303
         |||| ...|||||   ...|||...|||||...   |||...||||||.
     202 GATCAGATACGTCTTTAACTTGGCCGAGTTGTCAGCTTGTTCCCATGG 251

     303 spCysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeu... 318
         ::   :::  |||||||||   |||  :::::.|||   ...|||
     252 AGTATGTCGGCATACCAGAGTGCATCAAACAAGTGATCAAGAGCTTAAT 301

     319 ...LysGluLysPro 322
         |||:::::::||
     302 GGAAAACAAGAACCA 316

seq_name: gb_est29:AI620433

seq_documentation_block:
LOCUS       AI620433      486 bp    mRNA            EST       21-APR-1999
DEFINITION  tu47b05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2254161 3'
            similar to TR:Q12992 Q12992 NTF2 ;, mRNA sequence.
ACCESSION   AI620433
VERSION     AI620433.1  GI:4629559
KEYWORDS    EST.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949357.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael T. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2254161"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr28 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101149-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      119 a    117 c    120 g    130 t

ORIGIN


alignment_scores:
            Quality:  372.00              Length:    150
              Ratio:  3.153                 Gaps:      0
 Percent Similarity:  78.667   Percent Identity: 45.333

alignment_block:
 US-09-174-937-5 x AI620433/rev   .

 Align seq 1/1  to reverse of: AI620433  from: 1  to: 486

     173 GlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 194
         ||||||  ||||||.:::|||    |||:::::|||||||:::
     486 GGGTACCCCGTGAAGGGCCTCAAGGGTCAATCCTTTTGGCTTCCTGTTA 437

     189 eLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuP 208
         :||||||                 |||'|||:'::::
     436 TCTACCAAAACCAAGAATCCCCACCACCCTTTGTGTCATGGAACACTGGT 387

     206 heLeuTyrValIleSerThrLeuGluLeuMetValAlaGluAspTyrMet 225
         ||   |||::::::::::||||||::: ':::::|||||||':::|'::::
     386 TTAGGTATATGGTGGGAACTCTGAAGCTGCTAGTAGCTGAAAATTACCTG 337

     223 IleValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGl 242
         :.'|| .'' : :||: :|||  |||  ::::::|||     ::
     336 CTTGTTCATTTGAGTGGAGGCACAAGCAGGGCCCAAGTTCCACCTTAAAG 287

     239 yTrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnL 258
         :|||.....:|||||:    :::|||||||||||||||||||||||||
     286 CTGGATACGTCAGTGTTACCGTACCCTGGATCGGCGGCTACGGAAAAACCT 237

     256 euLysSerPheIleIleValHisProSerTrpPheIleArgThrIleLeu 275
         ||.............||||||  ....|.....:::::::::::::::
     236 TGCGAGCCCTGGTGGTTGTCCATGCTACATGGTATGTGAAAGCATTTCTG 187

     273 AlaValThrArgProPheIleSerSerLysPheSerSerLysIleLysI 292
         |||:::  ||||||||||||||| ||||||||:::  |||'':::::
     186 GCACTGCTTCGGCCCTTCATCAGTTCCAAATTCAGCACGAAAAATCCGTTT 137

     289 rValAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspCysIleH 308
         .....:|||||...|||||...  ||||||  ..:::|  :::'
     136 TCTGGACAGCCTGGGGGAGCTGGCCCAACTTATATCCCTGGATCAAGTCC 87

     306 isIleProGluSerIleIleAsnIleAspLeuLysLeuLysGluLysPro 322
         |||||||||||: ' '  :  :  ||||.::  ||||||
      86 ACATCCCTGAAGCTGTCAGACAGTGATTTTCACACTTAAGAATAAAGCCT 37

seq_name: gb_est12.AA296475

seq_documentation_block:
LOCUS       AA296475      236 bp    mRNA            EST       18-APR-1997
DEFINITION  EST10971 Umbilical vein endothelial cells II Homo sapiens cDNA 5'
            end, mRNA sequence
ACCESSION   AA296475
VERSION     AA296475.1  GI:1948809
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Ma...-Wai,...,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.,Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Sandusky,D.M., Shi...,P.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
```

```
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,C.L., Kunsch,C., Murphy,J., Li,P., Melessner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillon,P.J., Fannon,M.P., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On Apr 14, 1998 this sequence version replaced gi:693143.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org:80/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..236
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):194265"
                     /db_xref="taxon:9606"
                     /clone_lib="Umbilical vein endothelial cells II"
                     /cell_type="endothelial cell"
                     /dev_stage="fetus"
                     /note="Organ: umbilical vein, Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       66 a      54 c      57 g      55 t       4 others
ORIGIN


alignment_scores:
            Quality:  366.00             Length:     78
              Ratio:  5.014                Gaps:      0
 Percent Similarity:  93.590   Percent Identity: 93.590

alignment_block:
 US-09-174-937-5 x AA296475   ..

 Align seq 1/1  to: AA296475  from: 1  to: 236

        2 GluGluGluThrGluPheLeuGluLeuGlyThrArgIleSerArgProAs 18
          |:|||||| |:|||:||  |||||||||| |  ||||||||||||||||
        1 GAGGAGGAGACAGAGTTCNTTGAGCTCGGAACTAGGATATCAAGACCAAA 50

       18 nGlyLeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyV 35
          ||| | | | ||:|||| | | |||:|  ||||||||||||||||||||
       51 TGGACTACTGTCAGAGGATGTAGGAATGGACATCCCCTTTGAAGAGGGCG 100

       35 alLeuSerProSerAlaAlaAspMetArgProGluProProAsnSerLeu 51
          |||| |||||||||||||| ||||||||||| ||||   ||||||
      101 TGCTGAGTCCCAGTGCTGCAGACATGAGGCCTAACCTCNTAATTCTNTG 150

       52 AspLeuAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIl 68
          | |||| |||||||| |||||||||| |  ||||||||||||||||||
      151 GTTCTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATNT 200

       68 eAsnLeuSerLeuAspGlnSerGluGlySerIle 79
          |||||| | || ||| |||||||| |
      201 CAATCTTTCTCTGGACCAAAGTGAAGGATCTATT 234

seq_name: gb_est9:AA067188

seq_documentation_block:
LOCUS       AA067199      452 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm38b12.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:523775 5' similar to TR:G558644 G558644 NIP2. ?, mRNA
            sequence.
ACCESSION   AA067188
VERSION     AA067188.1  GI:1564954
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Bietz,H.N., Chuque,T.,
            Geisel,S., Kucaba,T., Lay,M., Le,M., Martin,C., Mitchell,M.,
            Schellenberg,K., Stupp,M., Taylor,E., Thomas,M., Underwood,K.,
            Theising,B., Wylie,T., Dodson,R., Waterston,R. and Wilson,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1564953.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free from the I.M.A.G.E.
            Consortium (info@image.llnl.gov) - for distribution information.
            MGI:917623
            Seq primer: -28m13 revl PE 1 1rimA Amersham
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:523775"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="DH10B (Escherichia coli strain)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
                     Oligo dT; Site M2 cells; a melanoma; metastatic; derivative of
                     the K-1735 (mouse) melanoma; Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; 5' adaptor sequence: 5' AAGTTGGCACGAG
                     3' >3' adaptor sequence: 5' TTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      145 a      90 c      [illegible] g      104 t
ORIGIN


alignment_scores:
            Quality:  358.00             Length:     [illegible]
              Ratio:  3.934                Gaps:      [illegible]
 Percent Similarity:  80.091   Percent Identity: [illegible]

alignment_block:
 US-09-174-937-5 x AA067188   ..

 Align seq 1/1  to: AA067188  from: 1  to: 452

      224 ValTyrLeuAsnGlyAlaThrProArgArgArgMetArgProGlyGlyTr 24
          ||:  |||  |||||||||||||||||||||||||||||||||||||||||
        1 ATTTACTTAAAATGGTGCGACAACTGAAAAAAAATGGCGAAGGGGGATG 50

      240 pMetLysLysCysTyrGlnMetIleAspArgAsnLeuArgLysAsnLeuL 28
          |||||||||| | | | | |||                          |
       51 GCTCAGGAGATGCTACCAGCAAAATTAATAACTATTAAAAAAAAATTTA 100

      257 ysSerPheIleIleValHisProSerTrpIleArgGluHisAla 270
          |||  |||||||||||||||||||||||||||||||||||||||
      101 AGTCTCTAATCATTGTGCACCCCTTTTGTTTATCCGAATATTTTTTCT 150

      274 ValThrArgProPheIleSerSerAspLeuArgSerLysIleLysTyrVa 32
          |||||||||||||||||||||||||||||||||||||||||||||||||
      151 GTTACAAGACCATTTATCAGTTCAAAGCTAAGTCAAAATTAAAATATGT 200
```

```
     290 lAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspCysIleHisI 307
         ::: .':::!!! ' :::  '!':::!  !!:::   :::  |
     201 CTTTAACTTGGCCGAGCTGGCAGAGCTTGTTCCATGGAGTATGTCGGCA 250

     307 leProGluSerIle       .       .       .       .    311
         ||||||||   |||
     251 IACCAGAGTGCATCAAACAGIATGAAGAAGAAAAGTTTAAAAAGAGACAA 300

     312 ......IleAsnIleAspLeuLysLeuLysGluLysPro 322
               ::::::   :::!!:::   '! :::::!!!
     301 AAAAGAGTTGATCAAGAGCTTAATGGAAAACAAGAACCA 339

seq_name: gb_est9:AA068579

seq_documentation_block:
LOCUS       AA068579      443 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm45e06.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:524482 5' similar to TR:3558944 3558944 NIP2. , mRNA
            sequence.
ACCESSION   AA068579
VERSION     AA068579.1  GI:1575939
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802463.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box #501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318330
            Seq primer: -28m13 rev1 EI from Amersham
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1..443
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE 524482"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SCID (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      143 a     89 c    105 g    105 t
ORIGIN


alignment_scores:
            Quality:  358.00             Length:    113
              Ratio:  3.934                Gaps:      1
 Percent Similarity:  88.591   Percent Identity:  68.177

alignment_block:
```

```
US-09-174-937-5 x AA068579   ..

Align seq 1/1  to: AA068579  from: 1  to: 443

    224 ValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyLeuGlyIr 24.
        :::!!!!!!!!!!!!!!!!!!  !:' !:::: !!!!!:::!:!'!!!:.
      1 ATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGATG 50

    240 pMetLysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLeuL 257
        ! :.:. .:::!!!!!!!!!:!!!! !!!'!!!!!!!!!!'!'!'!'''.
     51 GCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGGAAAAACGTTA 11.

    257 ysSerPheIleIleValHisProSerTrpPheIleArgThrIleLeuAla 27.
        !!!!!:::!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
    101 AGTCTCTAATCATTGTGCACCCCTCTTGGTTTATCCGAACACGTTCTGGCT 15

    274 ValThrArgProPheIleSerSerLysPheSerSerLysIleLysTyrVa 29.
        !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
    151 GTTACACGACCATTTATCAGTTCAAATTTAGTTCAGAACATCATCAAGTT 2.

    290 lAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspCysIleHisI 307
        |   ':::!!!:::!!!!!!!!:::   '  :::!!!!!!!!:::   :::
    201 CTTTAACTTGGCCGAGCTGGCAGAGCTTGTTCCATGGAGTATGTCGGCA 250

    307 leProGluSerIle.........                .......... 311
        ||||||||   |||
    251 IACCAGAGTGCATCAAACAGIATGAAGAAGAAAAGTTTAAAAAGAGACAA 3..

    312 ......IleAsnIleAspLeuLysLeuLysGluLysPro 322
              :  :   :::!!!::   '!'    :'!!
    301 AAAAGAGTTGATCAAGAGCTTAATGGAAAACAAGAACCA 339

seq_name: gb_est1:T31983

seq_documentation_block:
LOCUS       T31983        281 bp    mRNA            EST       ( ***-**-95
DEFINITION  EST41923 Human Uterus Homo sapiens cDNA 5' ... similar to None,
            mRNA sequence.
ACCESSION   T31983
VERSION     T31983.1  GI:614081
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chin,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,T.A., Nayar,V.D.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Sandek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,K.T.R.,
            Weidman,J.F., Li,Y., Barbazuk,D.B., Cai,J., Cupples,M.A.,
            Coleman,T.A., Collins,F.S., Dimke,D., Feng,P., Ferrin,A.,
            Fischer,C., Hastings,C.S., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.I., Kunsch,C., Li,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,I., Wei,Y.-F., Wiggs,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC14305
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
```

```
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .281
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):102321"
                      /db_xref="taxon:9606"
                      /clone_lib="Human Uterus"
                      /note="Organ: uterus"
BASE COUNT       89 a     41 c     63 g     86 t      2 others
ORIGIN


alignment_scores:
            Quality:  354.50             Length:     94
              Ratio:  4.171                Gaps:      2
 Percent Similarity:  90.426   Percent Identity: 70.213

alignment_block:
 US-09-174-937-5 x T31983   ..

 Align seq 1/1  to: T31983  from: 1  to: 281

     159 IleGluProTyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTy 175
         |||||||||||||:::::|||||||||||||||                 ||||||||
       5 ATTGAACCCTATAAAAAAGTTATCAGCCATGGG    ......GGATATTA 45

     175 rGlyAspGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProA 192
         |   ||||||||||||||||||:::|||||||||||:::|||||||:::|||:
      46 TNGGGATGGATTAAAIGCCATTGTIGIGIIIGCTGTCTGTTTCATGCCTG 95

     192 spSerSerArgAlaAspTyrHisTyrValMetGluAsnLeuPheLeuTyr 208
         ::|||||...   ...|||  :|||:::|||:::|||||||||   |||
      96 AAAGTAGTCAGCCTAACTATAGATACCTGATGCAACAATCTTTTTAAATAT 145

     209 ValIleSerThrLeuGluLeuMetValAlaGluAspTyrMetIleValTy 225
         |||||:::|||||||||||:|:::|||||||||:::|||||||||||||
     146 GTTATTGGCACIIIGGAGCTATTAGTAGCAGAAAACTACATGATAGTTTA 195

     225 rLeuAsnGlyAlaThrProArgArg.ArgMetProGlyLeuGlyTrpMet 241
         ||||||||||||||   ||||||  :::|||||:::|||||||||:::
     196 TTTAAATGGTGCAACAACTCGAAGAAAAAATGCCCAGTCTGGGATGGCTC 245

     242 LysLysCysTyrGlnMetIleAspArgArg 251
         :::  |||||||||:::|||||||||||
     246 AGGNAATGTTATCAGCAAATTGATAGAAGG 275
```

```
OM of: US-09-174-937-2 to  N_Geneseq_36:*   out_format . pfs

Date. Feb 9, 2000 11.41 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09174937/runat_09022000_085108_13316/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -STAPT=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-174-937-2
Query length: 371
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 103.050000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
N_Geneseq_36:T38228   +   935.50 1942.92  1.3e-100  1386  ! Apoptosis-regulating protein
N_Geneseq_36:W01120   +   935.50 1942.92  1.3e-100  1386  ! Human Nip2 encoding cDNA. Poly
N_Geneseq_36:X39973   +   238.00  469.61  1.5e-18   1227  ! Gastric cancer associated gene
N_Geneseq_36:X39976   +   238.00  474.54  7.9e-19    816  ! Gastric cancer associated gene
N_Geneseq_36:T31993   +   114.00  192.89  0.0038     4080  ! Nonsense-mediated mRNA decay
N_Geneseq_36:T31994   +   114.00  199.85  0.0016     2295  ! Nonsense-mediated mRNA decay
N_Geneseq_36:T62259   +   101.50  172.37  0.0532     2704  ! Schizosaccharomyces pombe orig
N_Geneseq_36:T73226   +   101.50  172.37  0.0532     2704  ! S. pombe origin of replicatio
N_Geneseq_36:T28329   +    92.50  142.99  2.84       5346  ! Rat nestin gene. Diagnosis of
N_Geneseq_36:Q70447   +    92.50  135.19  6.07      11236  ! Rat nestin gene - its product
N_Geneseq_36:Q12261   +    91.50  136.84  5.07       8224  ! Versican gene. Versican protei
N_Geneseq_36:T93091   +    91.00  146.04  1.56       3523  ! Human transcription repressor
N_Geneseq_36:V52220   +    89.50  122.85 30.49      18436  ! Streptococcus pneumoniae genc
N_Geneseq_36:I29472   +    89.00  139.92  3.12       4116  ! Muramidase released protein ge
N_Geneseq_36:V40629   +    88.50  133.69  7.60       6318  ! Musca domestica voltage-sensit
N_Geneseq_36:V32900   +    88.50  142.02  2.61       3171  ! Solanum tuberosum leaf L-type
N_Geneseq_36:V82498   +    88.50  133.69  7.60       6318  ! Calcium permeable voltage sens
N_Geneseq_36:T34620   +    88.00  140.35  3.23       3337  ! P. vivax ESP-1 blood stage ant
N_Geneseq_36:X15174   +    88.00  140.35  3.23       3337  ! DNA encoding a secreted blood
N_Geneseq_36:I87157   +    87.50  145.41  1.96       9036  ! P. falciparum transmission bl
N_Geneseq_36:Q93525   +    87.00  134.60  6.76       4507  ! SABP gene. New erythrocyte bin
N_Geneseq_36:T30055   +    87.00  134.54  6.91       4728  ! S. pneumoniae detection probe,
N_Geneseq_36:T72888   +    87.00  134.60  6.76       4507  ! Sialic acid binding protein co
N_Geneseq_36:Q54386   +    86.50  105.22 290.70      46899  ! T. niveum Cyclosporin synthet
N_Geneseq_36:Q78276   +    86.50  135.07  6.36       3973  ! T. niveum alanine-racemase. Ne
N_Geneseq_36:Q78281   +    86.50  135.07  6.36       3973  ! T. niveum alanine-racemase. Ne
N_Geneseq_36:I62781   +    86.00  137.17  4.86       3060  ! Urea transporter polypeptide c
N_Geneseq_36:V40630   +    86.00  128.41 14.95       6315  ! Musca domestica voltage-sensit
N_Geneseq_36:V82499   +    86.00  128.41 14.95       6315  ! Calcium permeable voltage sens
N_Geneseq_36:X04639   +    86.00  137.47  4.68       2985  ! Nucleic acid encoding an enzym
N_Geneseq_36:V07261   +    85.50  106.99 236.33      34303  ! Adenoviral vector plasmid pBH
N_Geneseq_36:Q05752   +    85.00  132.95  8.35       3641  ! Microspore-specific clone L10
N_Geneseq_36:Q75956   +    85.00  128.32 15.12       5340  ! S. cerevisiae scaur2P gene. Ne
N_Geneseq_36:Q92656   +    85.00  136.37  5.39       2745  ! Virulence determinant rpoV gen
N_Geneseq_36:Q92657   +    85.00  136.37  5.39       2745  ! Virulence determinant rpoV gen
N_Geneseq_36:T40092   +    85.00  141.99  2.02       1724  ! M. tuberculosis RNA polymerase
N_Geneseq_36:T48966   +    85.00  132.95  8.35       3641  ! Brassica napus microspore-spec
N_Geneseq_36:T47442   +    85.00  125.28 20.35       6869  ! Saccharomyces cerevisiae aurel
N_Geneseq_36:V38109   +    85.00  136.37  5.39       2745  ! Mycobacterium bovis large ORF
N_Geneseq_36:V38109   +    85.00  136.37  5.39       2745  ! Mycobacterium bovis virulence
N_Geneseq_36:X32300   +    85.00  130.57 11.33       4434  ! Fumonosin-resistance protein e
N_Geneseq_36:W21203_12 +   84.50   90.68  1.9e+3      22 ((( ! Incubation (13 of 17
N_Geneseq_36:I67135   +    84.00  134.86  6.53       2611  ! Human PAC protein kinase a
N_Geneseq_36:I94812   +    84.00  134.86  6.53       2611  ! Human PAC protein kinase a
N_Geneseq_36:V71086   +    84.00  131.02  4.06       2184  ! Green fluorescent protein

seq_name: N_Geneseq_36:T38228

seq_documentation_block:
ID   T38228 standard; cDNA; 1386 BP.
AC   T38228;
DT   18-DEC-1996  (first entry)
DE   Apoptosis-regulating protein Nip2 cDNA.
KW   Apoptosis-regulating protein; Nip2; adenovirus E1B 19K protein;
KW   Bcl-2; oncogene; cell death; cancer; tumour; immune disorder;
KW   diagnosis; gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             214. .1158
FT                   /*tag= a
PN   EP-733706-A2.
PD   25-SEP-1996.
PF   21-MAR-1996; 104542.
PR   21-MAR-1995; US-408095.
PA   (UYSL-) UNIV ST LOUIS.
PI   Chinnadurai G;
DR   WPI; 96-427055/43.
DR   P-PSDB; W00998.
PT   Nucleic acids encoding apoptosis regulating proteins - useful for
PT   diagnosing and treating immune disorders, malignancies, etc.
PS   Claim 1; Page 39-41; 60pp; English.
CC   6 cDNA clones (T38227-32) respectively code for novel apoptosis-
CC   regulating proteins Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13
CC   (W01997-W01002).  Nip clones were isolated from human cDNA
CC   libraries using a yeast two-hybrid stategy with adenovirus E1B
CC   19K protein (W01019) as bait. Bip clones were isolated using
CC   Bcl-2 oncoprotein (W01018) as bait.  E1B 19K and Bcl-2 both protect
CC   cells against apoptosis.  Nip1, Nip2, Nip3 bind to 19K, Bip1A and
CC   Bip5 bind to 19K and Bcl-2, and Bip13 binds to Bcl-2 alone.  Nucleic
CC   acids can be used in diagnostic assays, to produce recombinant
CC   apoptosis-regulating proteins, or to regulate cell death by
CC   increasing or suppressing expression of an apoptosis blocking gene
CC   such as bcl-2.
SQ   Sequence   1386 BP;    428 A;    249 C;    336 G;    373 T;


alignment_scores:
            Quality:  935.50              Length:    371
              Ratio:  2.491                 Gaps:      0
 Percent Similarity:  77.011   Percent Identity:  52.291

alignment_block.
 US-09-174-937-2 x T38228   ..

 Align seq 1/1   to: T38228   from: 1   to: 1386

       1 GluAlaThrLeuArgMetGluAsnValAspAlaLysGluGlnProLeuAs 21
         ||:::  |||||||||||||||:::' ::::::  :  :  :  :::::
     199 GAATTCATGGTTGAGAATTGAAGTCTTCAATTTAAATAATGAGCAAGA 248

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuG 38
         |'||||||:::|||  ||'|'|'|' ' :::  ::::::
     249 TGAAGATTTTCCGATACCTTTACCAGGAGATGATGACTATTGAAGCGAATA 298

      39 lySerProValGluAspThrSerSerProProAspThrLeuAspMetAsn 54
         ...  ..... ..... ||||||| ||| ||
     299 TACTAGCTATAACTGGACCAGACGACACTGATTCCTTCATTAGAAATTAA 348

      55 GlyAlaHisArgLysArgLysThrLeuValLeuProGlnIleSerLeu 71
         |||  :::... ||||||  '''::: ! ':::  :::::::::
     349 GGA.  AATAAAGTGAGAAACAAATTAATGCCTTGAGCATTAGCTTGAG 395

      71 rLeuAspGlnSerGluGlySerLeuAsnAspAspPheSerLeuThrP 86
         |||||||  ||:::|'''||:::
```



seq_name: N_Geneseq_36:X04120

seq_documentation_block.
ID X04120 standard cDNA 1386 BP
AC X04120;
DT 13-APR-1999 (first entry)
DE Human Nip2 encoding cDNA.

KW Human; Nip1, Nip2, Nip3; Bip1A, Bip5, Bip13, adenovirus; cell death;
KW viral infection; Bcl-2; protooncogene; mutational analysis; apoptosis;
KW E1B 19K protein; cell survival regulation; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 214. .1158
FT /*tag= a
PN US5858678-A.
PD 12-JAN-1999
PF 21-MAR-1995; 408095.
PR 02-AUG-1994; US-284139.
PR 21-MAR-1995; US-408095.
PA (UYSL-) UNIV ST LOUIS.
PI Chinnadurai G;
DR WPI; 99-152099/13.
DR P-PSDB; W94324.
PT Polypeptides that bind to anti apoptotic proteins - useful for
PT protecting against cell death induced by viral infection and to
PT modulate response to physical and chemical stimuli
PS Claim 2; Column 51-54; 41pp; English.
CC The present invention describes (1) a method for regulating cell death,
CC comprising exposing an isolated cell to a polypeptide selected from
CC Nip1, Nip2, Nip3, Bip1A, Bip5 and Bip13; (2) a method for neutralising
CC the activity of the adenovirus E1B 19 kD protein, the Bcl 2 protein or
CC the BHRF-1 protein, comprising exposing an isolated cell to a
CC polypeptide as in (1), and (3) a method for detecting molecules that
CC bind to at least one polypeptide as in (2), comprising lysing cells,
CC exposing the lysate to the polypeptide and detecting any molecule-
CC polypeptide aggregates. The methods are useful for providing proteins
CC able to bind to other proteins known to regulate cell survival e.g. it
CC is known that E1B 19K protein provides a survival function similar to
CC the cellular protooncogene bcl 2 gene product which is able to block
CC apoptosis in haematopoietic B and T cells. The present sequence encodes
CC human Nip2, given in the present invention.
SQ Sequence 1386 BP, 428 A, 243 C, 336 G, 379 T;

alignment_scores:
Quality. 935.50 Length. 348
Ratio: 3.491 Gaps: 8
Percent Similarity: 77.011 Percent Identity: 52.299

alignment_block:
US-09-174-937-2 x X04120 ..

Align seq 1/1 to: X04120 from: 1 to: 1386





```
seq_name: N_Geneseq_36:X39873

seq_documentation_block:
ID   X39873 standard; DNA; 1227 BP.
AC   X39873;
DT   02-JUL-1999 (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, Ohno M, Ohta Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PI   New isolated cancer associated nucleic acids and polypeptides -
PI   isolated using sera from cancer patients, used to develop products
PI   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 579-580; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   1227 BP;    388 A;    270 C;    334 G;    254 T;

alignment_scores:
            Quality:  238.00              Length:    1 3
              Ratio:  3.013                 Gaps:      4
 Percent Similarity:  76.599   Percent Identity:  61.466

alignment_block:
 US-09-174-937-2 x X39873   ..

 Align seq 1/1  to: X39873  from: 1  to: 1227

      12 AsnValAspValLysGluGluTrp...Gln.AspGluAspLeuProArgP 27
      857 AACCAGGATATCAAGACCAAATGGACTACTGTCAGAGGATGTAG?AATGG 906

      27 roLeuProGluGluThrGlyVal?luLeuLeuGlySerProVal?luAsp 43
      907 ACATCCCCTTTGAAGAGGGTGTG...CTGAGTCCCAGTGCCTCCA?ACATG 955

      44 ThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLysAr 60
      956 AGGCCTGAACCTCCTAATTCTCTGGATCTTAA?ACACTGATCC.GGAG ....

      60 gLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerGlnG 77
      1004 AATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGAGCAAA?TCAAG 1053

      77 lySerLeuLeuSerAspAsp.PheLeuAspThrProAspAspLeuAspIl 93
      1054 GATCTATTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACAT 1103

      93 eAsnValAspAspIleGluThrProAspGluThrAspSerLeuGluPheL 110
      1104 CAATGTGGATGAACTTGATACCCTGGATGAAGAGAGATCTTTTTDA?TACC 1153

      110 euGly 111
      1154 CTGGC 1158

seq_name: N_Geneseq_36:X39876

seq_documentation_block:
ID   X39676 standard; DNA; 816 BP.
```

```
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis, research, treatment; human;
KW   breast cancer; colon cancer, gastric cancer, renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 581, 787pp, English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;    231 A;    191 C;    209 G;    184 T;


alignment_scores:
            Quality:  238.00              Length:    103
              Ratio:  3.013                 Gaps:      4
 Percent Similarity:  76.699   Percent Identity: 51.456

alignment_block:
 US-09-174-937-2 x X39876   ..

 Align seq 1/1  to: X39876  from: 1  to: 816

      12 AsnValAspValLysGluGluTrp...Gln.AspGluAspLeuProArgP 27
         |||   |||:::|||   :::|||       :::||||||:::
     446 AACCAGGAIAICAAGACCAAAIGGACIACIGICAGAGGATGTAGGAATGG 495

      27 roLeuProGluGluThrGlyValGluLeuLeuSerProValGluAsp 43
         :::|:   |||   |||||   |||       |||   ::::::
     496 ACAICCCCIIIGAAGAGGGCGIG...CTGATTTCAGTGCTGCAGACATG 542

      44 ThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLysAr 60
            :::|||||||||::|||:....|+    ...:|||   ..:||
     543 AGGCCTGAAGCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAG 592

      60 gLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerGluG 77
           |:|:::|||||:::|||||::|   ||||||||||||||||
     593 AAICAAGCTCACAGCCCCAAATAICAATCTTT:CTGGACCAAAGTGAAG 642

      77 lySerLeuLeuSerAspAsp.PheLeuAspThrProAspAspLeuAspIl 93
         |||:::: |||:|||||||    ||||:::::|||||||||:::::|||||
     643 GAICIAI:CICTCIGATGATAACTTTGGA:ACT:CAGATGAAAITGACAT 692

      93 eAsnValAspAspIleGluThrProAspGluThrAspSerLeuGluPheL 110
         ||||||||||.:.....||||:|||||:::||||||:::|::::
     693 CAATGTGGATGAACTTGATACCCCCGATGAAGCAGATTCTTTTGAATACC 742

     110 euGly 111
           |||
     743 CTGGC 747

seq_name: N_Geneseq_36:T31993

seq_documentation_block:
ID   T31993 standard; DNA; 4080 BP.
AC   T31993;
DT   24-NOV-1996  (first entry)
DE   Nonsense-mediated mRNA decay 2 gene.
KW   Nonsense-mediated mRNA decay, NMD2; Upf1p, inhibition; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             362. .3633
FT                   /*tag= a
FT   intron          368. .480
FT                   /*tag= b
FT   misc_feature    368. .373
FT                   /*tag= c
FT                   /note= "5' splice site"
FT   misc_feature    426  432
FT                   /*tag= d
FT                   /note= "branchpoint"
FT   misc_feature    478. .480
FT                   /*tag= e
FT                   /note= "3' splice site"
FT   tata_signal     143. .149
FT                   /*tag= f
FT   protein_bind    164. .177
FT                   /*tag= g
FT                   /label= Upf1p_binding_consensus
PN   WO9622301-A1.
PD   25-JUL-1996.
PF   27-DEC-1995; U16930.
PR   20-JAN-1995; US-375300.
PA   (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PI   He F,  Jacobson AS;
DR   WPI; 96-354469/35.
DR   P-PSDB; W04896.
PT   Isolated nonsense-mediated mRNA decay gene and protein  used t.
PT   develop methods for inhibiting the decay pathway for producing
PT   heterologous or endogenous proteins
PS   Claim 1; Page 47-49; 73pp; English.
CC   The NMD2 gene (T31993) is named after its role in the Nonsense-
CC   Mediated mRNA Decay pathway.  The protein, Nmd2p, binds to Upf1p.
CC   A C-terminal fragment of the protein (T31994) also binds Upf1p and,
CC   when overexpressed in the host cell  the fragment inhibits the
CC   function of Upf1p, thereby inhibiting the nonsense-mediated
CC   mRNA decay pathway.
SQ   Sequence   4080 BP;    1449 A;    646 C;    800 G;    1185 T;


alignment_scores:
            Quality:  114.00              Length:    177
              Ratio:  1.295                 Gaps:      4
 Percent Similarity:  49.718   Percent Identity: 23.729

alignment_block:
 US-09-174-937-2 x T31993   ..

 Align seq 1/1  to: T31993  from: 1  to: 4080

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
         |||:::     :::|||  .::::       :::'''   ::::
    2971 GAGAGIGCTGTTCCIATCGAGTCAAICACCGAAGATGATGAGGATGAAGA 3020

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeu 40
         |.....          ::::::   |||||||:::|||||||
    3021 TGATGAA.................AACGACGATGGTGTCGATTTACTAG 3052
```

```
      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
         ||:::      :::          |||   |'|'''|||   :::
    3053 GAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAGCGCCA 3102

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
         ||'   |||:::  ::::::
    3103 GGAAAACATCAGGCAAAGCAA       ..        .......    ..... 3123

      71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
             |||:::|||||     ..       :::| '|||      |||
    3124 ....GACGAAAGTGAAGATGAAGACGATGAGGACGATGACGAGGATGATG 3169

      88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104
           |'|||||   '||   :::   ||||||   '''       ||||||
    3170 ACGATGACGATGACGATGATGATGATGATGAAGAAGAAGGCGATGAGGAT 3219

     105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121
         |||:::   ..          ..      ..       :||   ||:::|||
    3220 GATGATGAAGATGATGATGATGAGGATGATGATGATGAAGAAGAAGAAGA 3269

     121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
         ||::                                |||||         :
    3270 CAGC    ..... ..... ....      GACTCTGATTTGGAGT 3289

     138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154
         ::|||          |||   :::                 |||:::::::
    3290 ATGGTGGTGATCTTGACGCAGACAGAGATATTGAAATGAAACGAATGTAT 3339

     155 ArgThrVal....     IleIleGlyGluGlnGluHisArgIleAspLe 168
         :::          .     |||:::|||:::::   :::
    3340 GAAGAGTACGAGAGAAAACTAAAGGATGAGGAAGAAAGGAAAGCGGAAGA 3389

     168 uHisMetIleArgProTyrMetLysValVal 178
         ::::::   |||   ::::::|||:::::
    3390 AGAATTGGAAAGGCAATTTCAGAAAATGATG 3420

seq_name: N_Geneseq_36:I31994

seq_documentation_block:
ID   I31994 standard; DNA; 2295 BP.
AC   I31994;
DT   24-NOV-1996  (first entry)
DE   Nonsense-mediated mRNA decay 2 gene C-terminal (1089-3383).
KW   Nonsense-mediated mRNA decay; NMD2; Upf1p; inhibition; ss.
OS   Saccharomyces cerevisiae.
PN   WO9622301-A1.
PD   25-JUL-1996.
PF   27-DEC-1995; U16930.
PR   20-JAN-1995; US-375300.
PA   (UYMA-) UNIV MASSACHUSETTS MEDICAL CENT.
PI   He F,  Jacobson AS;
DR   WPI; 96-354469/35.
DR   P-PSDB; W01897.
PT   Isolated nonsense-mediated mRNA decay gene and protein - used to
PT   develop methods for inhibiting the decay pathway for producing
PT   heterologous or endogenous proteins
PS   Claim 2, Page 52-54, 73pp; English.
CC   The NMD2 gene (I31993) is named after its role in the Nonsense-
CC   Mediated mRNA Decay pathway.  The protein, Nmd2p, binds to Upf1p.
CC   A C-terminal fragment of the protein (I31994) also binds Upf1p and,
CC   when overexpressed in the host cell, the fragment inhibits the
CC   function of Upf1p, thereby inhibiting the nonsense-mediated
CC   mRNA decay pathway.
SQ   Sequence  2295 BP;    891 A;    359 C;    447 G;    598 T;


alignment_scores:
            Quality:  114.00            Length:    177
              Ratio:  1.295               Gaps:      4
 Percent Similarity:  49.719   Percent Identity: 23.729

alignment_block:
```

```
US-09-174-937-2 x I31994   ..

Align seq 1/1  to: I31994  from: 1  to: 2295

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
         |||||||||||||   ::: |||||||          :::      :::
    1522 GAGAGTGCTTTTTTATTTAATTAATTTTTTTTTAATATAAAAAAAAAA 1571

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuG 38
         |:::::                    ::::::   '''  :::
    1572 TGATGAA .....          ..  AACGACGATGGTGTCGATTTACTAG 1613

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
         ||:::      :::          |   |   |   |   :::
    1604 GAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAGCGCCA 1653

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
         |||   ||:::  ::::::
    1654 GGAAAACATCAGGCAAAGCAA.............................. 1674

      71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
           |'|:::|||||'   :::      ::::  '''
    1675 ....GACGAAAGTGAAGATGAAGACGATGAGGACGATGACGAGGATGATG 1724

      88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104
         |'|||||   '||   :::   '|'||      |   |  :::::
    1721 ACGATGACGATGACGATGATGATGATGATGAAGAAGAAGGCGATGAGGAT 1770

     105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121
         |||:::   :::          :::   ::::::   |   |  :::::
    1771 GATGATGAAGATGATGATGATGAGGATGATGATGATGAAGAAGAAGAAGA 1820

     121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
         ||::                                |           :
    1821 CAGC    ...          ................GACTCTGATTTGGAGT 1867

     138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154
         ::|||          |||'   :::                 ::::::
    1841 ATGGTGGTGATCTTGACGCAGACAGAGATATTGAAATGAAACGAATGTAT 1891

     155 ArgThrVal.........IleIleGlyGluGlnGluHisArgIleAspLe 168
         :::          :::       | |||| |:::::::   :::
    1891 GAAGAGTACGAGAGAAAACTAAAGGATGACGAACAAAGCAAACGCGAACA 1940

     168 uHisMetIleArgProTyrMetLysValVal 178
         ::::::   |||   :::::: | ::::::
    1941 AGAATTGGAAAGGCAATTTCAGAAAATGATG 1971

seq_name: N_Geneseq_36:T62359

seq_documentation_block:
ID   T62359 standard; cDNA; 2504 BP.
AC   T62359;
DT   23-JUL-1997  (first entry)
DE   Schizosaccharomyces pombe origin of replication complex ORC1 gene.
KW   Origin of replication complex; ORC; gene therapy; cancer;
KW   neoplasia; inflammation; hypersensitivity; ds.
OS   Schizosaccharomyces pombe.
FH   Key             Location/Qualifiers
FT   cds             86. .2209
FT                   /*tag=  a
FT   misc_difference 2350. .2379
FT                   /*tag=  b
FT                   /note= "degenerate portion of the sequence"
PN   WO9640977-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U09403.
PR   07-JUN-1995; US-484105.
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
PI   Bell SP,  Foss M,  Horskowitz I,  Kobayashi R,  Laurenson P;
PI   Li JJ,  McNally FJ,  Rine J,  Stillman BW;
DR   WPI; 97-052354/05.
```

```
DR   P-PSDB, W14137.
PT   Nucleic acid encoding origin of replication complex (ORC) protein -
PT   useful to screen for lead pharmaceuticals capable of disrupting ORC
PT   protein function, and inhibiting cell growth
PS   Disclosure; Page 22-24; 57pp; English.
CC   Isolated cDNA clones (T62358-63) respectively encode origin of
CC   replication (ORC) proteins (W14136-41) from Kluyveromyces lactis,
CC   Schizosaccharomyces pombe, human (IBC1), Arabidopsis thaliana,
CC   Caenorhabditis elegans and human (ORC1). The S. pombe ORC1 clone
CC   was obtd. by PCR amplification using primers based on Saccharomyces
CC   cerevisiae ORC1 and SIR3 sequences. The isolated nucleic acids can
CC   be utilised in the prodn. of ORC polypeptides, to design probes and
CC   primers for the detection and amplification of ORC genes, and in
CC   gene therapy applns., e.g. antisense oligonucleotides capable of
CC   inhibiting the intracellular expression of a targetted ORC
CC   transcript.
SQ   Sequence   2504 BP;   785 A;   439 C;   499 G;   759 T;

alignment_scores:
           Quality:  101.50            Length:    341
             Ratio:  0.672               Gaps:     16
 Percent Similarity:  44.282   Percent Identity:  19.941

alignment_block:
 US-09-174-937-2 x T62359

 Align seq 1/1 to: T62359 from: 1 to: 2504

      59 LysArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSe 75
      92 AGAAGAAAGTCATTGAGGAGTCAACTATTAATTAACGGCATTGATAAA.. 139

      75 rGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuA 92
     140 .......AGTCTGCTATCTGATGACAGCGCTGACAGTTCTGAT....... 175

      92 spIleAsnValAspAspIleGlnThrProAspGluThrAspSerLeuGlu 108
     176 ..................................ATTGACGAAGAGGAA 190

     109 PheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProValAl 125
     191 GTTTACGGTGTTTGGACTGAAGAGCGCTTTCAAAAAGAGGCTCGACGTTC 240

     125 aThrAlaLysAsnMetPro                      GlyAspS 134
     241 TTATTACAGATCTTTAAAGAAAAACGATGTAATATATCGCGTTGGAGAT. 289

     134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 150
     290 ............GATATTACTGTACATGATGGAGACTCAAGC... 319

     151 GlyArgLeuTrpArgThrValIleIleGlyGlnGlnGluHisArgIleAs 167
     320 ......TTTTATCTGGGGGTAATTTGTAAATTGTACGAAAAAGCAATTGA 363

     167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrT 184
     364 TAAGCATTCTGGAAAGAAATATGTTGAAGCAATTGGTATAGTCGAGCTT 413

     184 yrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuPro 200
     414 ATGCTAAGAGAATGGAA......................ATTAAACCT 439

     201 AspSerSerLeuProAspTyrHis...........TyrIleMet..... 211
     440 GAATATTTGTTGCCAGACCGGCATATAAATGAGGTGTACGTTTCTTGTGG 489

     212 .......GluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuV 226
     490 CCGGGATGAAAACCTGACTTCATGTATAATAGATCATTGTAATGTCTACT 539

     226 alAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArg 242
     540 CTGAAGCAGAGTTTTTTTCAAAATTTCCCGTTGGAATTCCTACAAAACGA 589

     243 ArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspAr 259
     590 AAAGAT.............TTGTTTCCTTGT................. 607

     259 gArgLeuArgLysAsnLeuLysSerIleIleIleValHisProSerTrpP 276
     608 ..........................................AACTTCT 614

     276 heIleArgThrValLeuAlaIleSerArgProPheIleSerValLysPhe 292
     615 TTATCCGACGCGGTGTACACTTGAAA...................... 640

     293 IleAsnLysIle..........................GlnTyrValHisSerLe 302
     641 GTGAACAAATACACAGAACCTCTCGATTGGTCTTATTATGCTCATAATCT 690

     302 uGluAspLeuGluGlnLeuIleProMetGluHisValGlnIleProAspC 319
     691 TGAAAGGATAGAA..........................GATC 707

     319 ysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArgGluSer... 334
     708 TTTTGGTTGAGATGGAAGAAAATTTCCGACCAACTAAAAAGAAATTTGGT 757

     335 ...........AlaArgProGlnProGluPheValLeuProArgSerGl 347
     758 TCTAGAGGTCGTGGTCGCCCTCGT   AAATATCCTTTACCAAATGTCGA 804

     347 uGluLysProGluValAlaProValGluAsnArgSerAlaLeu......V 362
     805 AAGCAAAGAAAGCAGTTCCAAAGTTAACTCTAAGGATGAAAATTTTGATT 854

     362 alSerGluAspGlnGluThrSer 369
     855 TACAAGATGATAGTGAATCTTCA 877

seq_name: N_Geneseq_36:T73286

seq_documentation_block:
ID   T73286 standard; cDNA; 2504 BP.
AC   T73286;
DT   12-SEP-1997  (first entry)
DE   S. pombe origin of replication complex protein 1 gene.
KW   Origin of replication complex; ORC; yeast; Schizosaccharomyces pombe;
KW   chromatography; peptide sequencing; primer; amplification; PCR; genome;
KW   polymerase chain reaction; open reading frame; cell growth; cancer;
KW   infection; inflammation; hypersensitivity; ds.
OS   Schizosaccharomyces pombe
FH   Key             Location/Qualifiers
FT   cds             86..2203
FT                   /*tag= a
FT                   /product= ORC1
FT                   /note= "stop codon is not specified at the end of the
FT                   CDS"
PN   US5614618-A.
PD   25-MAR-1997.
PF   16-DEC-1993; 168479.
PR   16-DEC-1993; US-168479.
PR   07-JUN-1995; US-484106.
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
PI   Bell SP,  Foss M,  Gavin K,  Herskowitz I,  Hidaka M;
PI   Kobayashi R,  Laurenson P,  Li J,  McNally FJ,  Rine J;
PI   Stillman BW;
DR   WPI; 97-201534/18.
DR   P-PSDB, W22231.
PT   Nucleic acids encoding origin of replication complex proteins - used
PT   for screening for lead cpds. for therapy or diagnosis of disease
PT   associated with undesirable cell growth
```

```
PS   Claim 4; Column 65-68; 53pp; English.
CC   This is the nucleotide sequence encoding the origin of replication
CC   complex protein 1 (ORC1) from the yeast Schizosaccharomyces pombe.
CC   The sequence was isolated using primers based on amino acid sequence
CC   conserved between the ORC1 and SIR3 proteins from Saccharomyces
CC   cerevisiae.  The amplified fragment was then used for low stringency
CC   DNA hybridisation to obtain the S. pombe ORC1 gene sequence.  The ORC
CC   proteins (W22224-39) can be used to screen chemical libraries to
CC   identify lead compounds useful in treatment and diagnosis of undesired
CC   cell growth, e.g. cancer, infections, inflammation and hypersensitivity.
SQ   Sequence   2504 BP;    785 A;    499 C;    499 G;    759 T;


alignment_scores:
            Quality:  101.50              Length:    341
              Ratio:  0.672                Gaps:     16
 Percent Similarity:  44.282   Percent Identity: 19.941

alignment_block:
 US-09-174-937-2 x T73286   ..

 Align seq 1/1  to. T73286  from. 1  to. 2504

      59 LysArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSe 75
         :::|||||:::|||    .           .  :|||||:
      92 AGAAGAAAGTCATTGAGGAGTCAACTATTAATTAACGGCATTGATAAA.. 139

      75 rGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuA 92
               |||||||||||||||||     ||:::   |||
     140        AGTCTGCTATCTGATGACAGCGCTGACAGTTCTGAT        175

      92 spIleAsnValAspAspIleGluThrProAspGluThrAspSerLeuGlu 108
                                              |||:::   |||
     176 ................................ATTGACGAAGAGGAA 190

     109 PheLeuGlyAsnGlyAsnGluLeuGluTrpGlnAspAspThrProValAl 125
            ||      ::|||'      ::::::   ::::::      ::
     191 GTTTACGGTGTTTGGACTGAAGAGCCCTTTCAAAAAGAGGCTGGACGTTC 240

     125 aThrAlaLysAsnMetPro.......................GlyAspS 134
         :    ::::::::::                           ||||||
     241 TTATTACAGATCTTTAAAGAAAAACGATGTAATATATCGCGTTGGAGAT. 289

     134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 150
                     |||   |||          ||| .. ..
     290 ..............GATATTACTGTACATGATGGAGACTCAAGC... 319

     151 GlyArgLeuTrpArgThrValIleIleGlyGlnGluHisArgIleAs 167
            ::::::      ||||||           |||      |||||
     320 ......TTTTATCTGGGGGTAATTTGTAAATTGTACGAAAAAGCAATTGA 363

     167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrT 184
         |   '''|      :::   |'|::::::::::::   ::::::     |
     364 TAAGCATTCTGGAAAGAAATATGTTGAAGCAATTTGGTATAGTCGAGCTT 413

     184 yrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuPro 200
         |:::::   ::::::                    :::   |||
     414 ATGCTAAGAGAATGGAA.....................ATTAAACCT 439

     201 AspSerSerLeuProAspTyrHis............TyrIleMet..... 211
         :::    ''|'::::|   ''                |||...
     440 GAATATTTGTTGCCAGACCGGCATATAAATGAGTGTACGTTTCTTGTGG 489

     212 ........GluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuV 226
                 ||||||||||         :::|'::::      ::::::
     490 CGGGGATGAAAACCTGACTTCATGTATAATAGAGCATTGTAATGTCTACT 539

     226 alAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArg 242
             :::::::::    ''          |       ...|||
     540 CTGAAGCAGAGTTTTTTTCAAAATTTCCGTTGAATTCCTACAAAACGA 589

     243 ArgMetProGlyIleGlyTrpLeuLysLysGly.TyrGlnMetIleAspAr 259
         :::                 ''

     590 AAAGAT..............TTGTTTGTTGT.................. 611

     259 gArgLeuArgLysAsnLeuLysS.tr.IleI..ValH..rP..rTtt! 271
                                                 ::::::
     608 ..................................AATTTT 614

     272 heIleArgThrValLeuAlaIleserArgProIeIleserVa.LysPhe 287
         | ' ' '|      :::   ::::::
     615 TTATCCGACGCGGTGTACACTTCGAAA..................... 654

     288 IleAsnLysIle.....................GlnTyrValHisSerLe 302
         ::: |-| |                       .:::   :::
     641 GTGAACAAATACACAGAACCTCTTGATTGGTCTTATTATTTTCATAATTT 690

     303 uGluAspLeuGluGlnLeuIleProMetGluHisValGlnIleProAspL 319
         |''|'   :::|||                                   '
     691 TGAAAGGATAGAA.....................................GATT 727

     319 ysValLeuGlnTyrGluGluGlnArgLeuLysAlaArgArgGlnSer... 334
           :::::::::   | ' ' |:::          :::::::::::::
     728 TTTTGGTTGAGATGGAAGAAAATTTGCGACGAATAAAAAGAAATTTGT 777

     335 .............AlaArgProGlnProGluPheValIleProArgSer 347
                     ::: ' '|::::   ::::::      :::
     778 TCTAGAGGTCGTGGTCGCCTTCGT...AAATATTCTTTACTAAATTGTGA 824

     347 uGluLysProGluValAlaProValGlnAspArgSerAlaLe.......V 361
         ':::'|   :::   :::          :::::::::::::       :
     805 AAGCAAAGAAAGCAGTTCCAAAGTTAACTCTAAGGATGAAAATTTTGATT 854

     362 alSerGluAspGlnGluThrSer 374
         .....|||...|'|...||
     855 TACAAGATGATAGTGAATCTTCA 877


seq_name: N_Geneseq_36:Q28398

seq_documentation_block:
ID   Q28398 standard; DNA; 5946 BP.
AC   Q28398;
DT   15-FEB-1993  (first entry)
DE   Rat nestin gene.
KW   Intermediate filament; central nervous system; brain tumour;
KW   neurofilament; ss.
OS   Rattus norvegicus.
FH   Key             Location/Qualifiers
FT   cds             128. .5545
FT                   /*tag=  a
FT                   /product= nestin
FT                   /note= "128 could be 161"
FT   exon            1. .913
FT                   /*tag=  b
FT                   /number= 1
FT   exon            914. .1038
FT                   /*tag=  c
FT                   /number= 2
FT   exon            1039. .1110
FT                   /*tag=  d
FT                   /number= 3
FT   exon            1111. .5946
FT                   /*tag=  e
FT                   /number= 4
FT   polya_signal    5924. .5929
FT                   /*tag=  f
PN   WO9214821-A.
PD   03-SEP-1992.
PF   21-FEB-1992; 92WO-US01375.
PR   22-FEB-1991; US-660412.
PA   (MASS ) MASSACHUSETTS INST TECHNOLOGY.
PI   Lendahl U, McKay RDG;
DR   WPI; 92-316175/38.
DR   P-PSDB; R27204.
PT   Diagnosis of pre-disposition to brain tumours - using DNA
PT   encoding nestin which distinguishes neural stem cells from
```

```
PT   cells from neuronal, glial and muscle cells
PS   Claim 3; Fig 1; 53pp; English.
CC   A lambda gt11 cDNA expression library was constructed from poly(A)+
CC   RNA from CNS of day 15 rat embryos. The library was screened with
CC   MAb Rat 401. cDNA from the most immunopositive clone was used as
CC   a probe to isolate additional clones from a lambda gt10 library
CC   from rat E15 CNS. Hybridising clones were sequenced; the remainder
CC   of the nestin gene was obtained from genomic sequences. There are
CC   two potential initiation codons at positions 128 and 160. Sequence
CC   comparison of cDNA and genomic clones reveals three introns at
CC   positions 912, 1038 and 1111. The sequence has significant
CC   similarities to the five classes of intermediate filaments.
CC   The coding sequence can be used to develop probes for
CC   detecting predisposition to brain tumours; the expression of nestin
CC   distinguishes neural multi potential stem cells and brain tumour
CC   cells from the more differentiated neural cell types (e.g. neuronal,
CC   glial and muscle cells) of the mammalian brain.
CC   NOTE: the rat nestin coding sequence is printed in the patent
CC   specification but is illegible; this sequence has been obtained from
CC   the EMBL record (Accession number M34384).
CC   See also Q28399 for the human nestin gene.
SQ   Sequence   5946 BP;    1600 A;    1411 C;    1671 G;    1074 T;


alignment_scores:
            Quality:   92.50            Length:    238
              Ratio:  0.881              Gaps:     13
 Percent Similarity:  44.118   Percent Identity: 24.370

alignment_block.
 US-09-174-937-2 x Q28399  ..

 Align seq 1/1  to  Q28398  from  1  to  5946

       4 ThrGluAlaThrLeuArgMet.........GlnAsnValAspValLys.. 16
         |||:::  |:||||||||          || :.....  :|||:::
    3119 ACCCAACAAACACTGAGGGCTGTAGGGAATGAGCAGATGGCAGTGAGCCC 3168

      17 .GluGluTrpGlnAspGluAspLeuProArgProLeuPro..... 29
          ||'     |||  :::|||||||:::|:'|||
    3169 ACCAGAAAAGGTGGATCCAGAGTTACCGAAGCCTCTTGGAAATGACCAGG 3218

      29                                          . . . . 29

    3219 AAATAGCTAGATCTCTTGGAAAAGAGAATCAAGAGTCACTAGTGTCACTG 3268

      30 GluGluThrGlyValGluLeuLeuGlySerProValGluAspThrSerSe 46
         :::|||  |||:::|||  :::  ||         :::      ::
    3269 AAAGAAAAAGGTATAGAGACAGTGAAGTCTTTAGAAACAGAGATCATAGA 3318

      46 rProProAsnThrLeuAsnPheAsnGlyAlaHisArgLys.        . 59
         :|||  :::|||  :::  :::      :::||||||
    3319 ACCACTGGAGACTGCAGAAGAGGACCTGGAAACAAGGAAGTCTATAGATA 3368

      60                            .  ..ArgLysThrLeuValAla 65
                                       . |:::|||:::
    3369 CTCAGGAGCCAATTGTGGTCTACTGAAGTGGCTAGAGAGACAGTAGAACCT 3418

      66 ProGluIleAsnIleSerLeuAspGlnSerGluGlySerLeuLeuSer.. 81
         |'|||          |||:::      ||||||||  |||
    3419 CCAGAA..............GATGAGCCCCCAGGATCGCTAGGGTCTGT 3453

      82  AspAspPheLeuAspThrProAspAspLeu         ...     . 91
        .|:::      ...|:|     ...|:|
    3454 GGATGAGAACCGAGAGACACTGACATCCCTTGAAAAGGAGAGTCAAGAAC 3503

      92                            AspIleAsnValAspAsp..... 97
                                    :..   ...|||:::|||
    3504 TGAGCTCTCTGGGCAAGTGGAACGTAGAGACCAAGGTAGAGGACAGTCAG 3553

      98 .....................IleGluThrProAspGluThrAspSe 106
                              ..... ..... ...|||
    3554 CAGTGCCTGCAAGTAGAAGAGGGTCTGCAGGAG-AACAGCAGCAAGAGTC 3602

     106 rLeu.............GluPheLeuGlyAsnGlyAsnGluLeuGluT 118
         ||||             |||:::  :::::|||||:::   :::|
    3603 TCTGAGAGAGGTGAAGCAGGAGCTGCCTAGCTCTGGAAATCAACAGCGGT 3652

     118 rpGluAsp .                    AspThrProValAla 125
         ||||||||                     :::::|':::::
    3653 GGGAGGATGTGGTGGAGGGCAAAGCAGTGGGTCAGGAAGCACCTCTGGCA 3702

     126 ThrAlaLysAsnMetProGlyAspSerAlaAspLeuPhe.....GlyAs 140
         |||···              |||::::|||::|||        ||||:
    3703 ACCACAGGAGTGGGAACTGAGGATAAGGCAGAGTTGCATCTGAGGGGGCA 3752

     140 pGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrV 157
         :|||     |||···  :··|||||···|||··:|||   :::
    3753 AGGTGGAGAGGAAGAA .GCTGCAGCAGAGGGAGAGCTGTTGCAGGAT. 3799

     157 alIleIleGlyGlu 161
           |||::|||||
    3800 ..ATTGTGGGGGAG 3811

seq_name. N_Geneseq_36.Q70447

seq_documentation_block:
ID   Q70447 standard; DNA; 11236 BP.
AC   Q70447;
DT   21-MAR-1995  (first entry)
DE   Rat nestin gene - its product is useful to identify brain tumours.
KW   nestin gene; brain tumour; neoplastic cells; glial; neuronal;
KW   muscle; neural multipotential stem cell; mammalian brain; detection;
KW   diagnosis, medulloblastoma, glioblastoma, oligodendroglioma; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   misc_feature    2463
FT                   /*tag= a
FT                   /note= "start of primary transcript"
FT   cds             2589. .10821
FT                   /*tag= b
FT                   /product= Nestin protein
FT   misc_feature    3089. .3090
FT                   /*tag= c
FT                   /transl_except= pos:3087. .3091, aa:His, Arg
FT                   /note= "sequence should be CAY CGG, ie. Y or G has
FT                   been deleted in the sequence given"
FT   misc_feature    3104
FT                   /*tag= d
FT                   /note= "apparent inclusion of a nucleotide"
FT   intron          3375. .4339
FT                   /*tag= e
FT   exon            4340. .4464
FT                   /*tag= f
FT   intron          4465. .6119
FT                   /*tag= g
FT   exon            6120. .6193
FT                   /*tag= h
FT   intron          6194. .6388
FT                   /*tag= i
FT   exon            6389. .10821
FT                   /*tag= j
PN   US5338839-A.
PD   16-AUG-1994
PF   12-APR-1988; 180548.
PR   12-APR-1988; US-180548.
PR   02-JUN-1988, US-201762.
PR   25-OCT-1990; US-603803.
PR   22-FEB-1991; US-660412.
PR   19-MAR-1992; US-853913.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI   Lendahl U, Mckay RDG;
DR   WPI; 94-263332/32.
DR   P-PSDB; R60126.
PT   Nucleotide and protein sequences for human and rat nestin -
PI   distinguishes neural multipotential stem cells and brain tumour
```

```
PT   cells from more differentiated cell types, for use in the
PT   diagnosis of brain tumours
PS   Claim 1; Column 25-34, 45pp; English.
CC   Q70447 is the rat nestin gene encoding nestin protein (R60126).
CC   Nestin protein expression distinguishes neural multipotential stem
CC   cells and brain tumour cells from the more differentiated neural
CC   cell types (eg., neuronal, glial and muscle cells of the adult brian).
CC   The nestin protein can be used in diagnosing tumours of the brain,
CC   such as medulloblastomas, gliablastomas and oligodendroglioma.
CC   (See also Q70448).
SQ   Sequence   11236 BP;    2876 A;    2678 C;    3258 G;    2424 T;

alignment_scores:
          Quality:   92.50              Length:    238
            Ratio:   0.881                Gaps:     13
 Percent Similarity:  44.118    Percent Identity: 24.370

alignment_block:
 US-09-174-937-2 x Q70447   ..

 Align seg 1/1  to: Q70447  from: 1  to: 11236

      4 ThrGluAlaThrLeuArgMet.  ......GluAsnValAspValLys . 16
        ||||::   |||||||||       |||            |||:::
   8395 ACCCAACAAACACTGAGGGCIGIAGGGAAIGAGCAGATGGCAGTGAGCCC 8444

     17 .GluGluTrpGlnAspGluAspLeuProArgProLeuPro.......... 29
             |||       |||    :::||||||||:::|:::|||
   8445 ACCAGAAAAGGTGGATCCAGAGTTACCGAAGCCTCTTGGAAATGACCAGG 8494

     29 ..                                      ........ 29

   8495 AAATAGCTAGATCTCTTGGAAAAGAGAATCAACAGTCACTAGTGTCACTG 8544

     30 GluGluThrGlyValGluLeuLeuGlySerProValGluAspThrSerSc 46
        :::|||   |||...|||     ·    :||        :::       ··
   8545 AAAGAAAAAGGTATAGAGACAGTGAAGTCTTTAGAAACAGAGATCATAGA 8594

     46 rProProAsnThrLeuAsnPheAsnGlyAlaHisArgLys.......... 59
        ·|||      |||    :::    :          ·  ||||||
   8595 ACCACTGGAGACTGCAGAAGAGGACCTGGAAAGAAGGAAGTCTATAGATA 8644

     60                        .......    .ArgLysThrLeuValAla 65
                                            |:::|||:::
   8645 CTCAGGAGCCATTGTGGTCTACTGAAGTGGCTAGAGAGACAGTAGAACCT 8694

     66 ProGluIleAsnIleSerLeuAspGlnSerGluGlySerLeuLeuSer.. 81
        ||||||                 |||···      ||||||||    |||
   8695 CCAGAA.               ...GATGAGCCCCCAGGATCGCTAGGGTCTGT 8729

     82 .AspAspPheLeuAspThrProAspAspLeu................... 91
         |||···        ···|||       ··:|||
   8730 GGATGAGAACCGAGAGACACTGACATCCCTTGAAAAGGAGAGTCAAGAAC 8779

     92 .                   ....AspIleAsnValAspAsp...... 97
                                 ···    ··|||...|||
   8780 TGAGCTCTCTGGGCAAGTGGAACGTAGAGACCAGGGTAGAGGACAGTCAG 8829

     98                  ..........IleGluThrProAspGluThrAspSe 106
                                    ::::::      ::::::    :::||
   8830 CAGTGCCTGCAAGTAGAAGAGGGTCTGCAGGACCAACAGCACCAAGAGTC 8879

    106 rLeu....            GluPheLeuGlyAsnGlyAsnGluLeuGluT 118
        ||||                ||||::   :::::||||||:::   :::|
   8880 TCTGAGAGAGGTGAAGCAGGAGCTGCCTAGCTCTGGAAATCAACAGCGGT 8929

    118 rpGluAsp...           .....          AspThrProValAla 125
        |||||||||                            ::····|||··:|||
   8930 GGGAGGATGTGGTGGAGGGCAAAGCAGTGGGTAGGAAGCACCTCTGGCA 8979

    126 ThrAlaLysAsnMetProGlyAspSerAlaAspLeuPhe......GlyAs 140
        |||:::               |||:::||||:::|||            |||::

   8980 ACCACAGGAGTGGGAACTGAGGATAAGGCAGAGTTGTATCTGAGGGGGCA 9029

    140 pGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrV 157
         :|||       |||:::     :::       :::     :::        :::
   9030 AGGTGGAGAGGAAGAA...GCTGCAGCAGAGGGAGAGCTGTTGCAGGAT. 9075

    157 alIleIleGlyGlu 161
          |||::::||||
   9076 ..ATTGTGGGGGAG 9087

seq_name. N_Geneseq_36.Q12261

seq_documentation_block:
ID   Q12261 standard; cDNA; 8224 BP.
AC   Q12261;
DT   11-SEP-1991  (first entry)
DE   Versican gene.
KW   Fibroblast proteoglycan; hyaluronic acid binding;
KW   tissue reconstruction; ss.
OS   Homo sapiens.
FH   Key                 Location/Qualifiers
FT   CDS                 267..7496
FT                       /*tag= a
FT   signal_peptide      267..326
FT                       /*tag= b
FT   mat_peptide         327..7493
FT                       /*tag= c
PN   WO9108230-A.
PD   13-JUN-1991.
PF   27-NOV-1990; U06897.
PR   27-NOV-1989; US-441179.
PA   (JOLL-) LA JOLLA CANCER RES.
PI   Ruoslahti E;
DR   WPI; 91-193156/26.
PT   Versican proteoglycan and nucleic acid - having hyaluronic acid
PT   binding region and used in e.g., tissue reconstruction
PS   Claim 1; Fig 1; 27pp; English.
CC   The DNA encodes versican, a new fibroblast proteoglycan. The
CC   sequence was obtd. from a clone isolated from a human placental
CC   library.  The encoded protein comprises a signal peptide, N-term-
CC   inal hyaluronic acid domain, C-terminal EGF like domain, lectin-
CC   like sequence and complement regulatory protein-like domain. The
CC   DNA can be used to express versican in transformed hosts. The
CC   recombinant protein can be used to determine the presence of hyal-
CC   uronic acid, and as a vehicle to bring other molecules into con-
CC   tact with hyaluronic acid.  Fragments of the DNA can also be used
CC   as probes.  Antibodies can be prepared against te N-terminal or
CC   glycosaminoglycan attachment domain.
SQ   Sequence   8224 BP;    2524 A;    1843 C;    1796 G;    2061 T;

alignment_scores:
          Quality:   91.50              Length:    443
            Ratio:   0.446                Gaps:     21
 Percent Similarity:  45.657    Percent Identity: 19.822

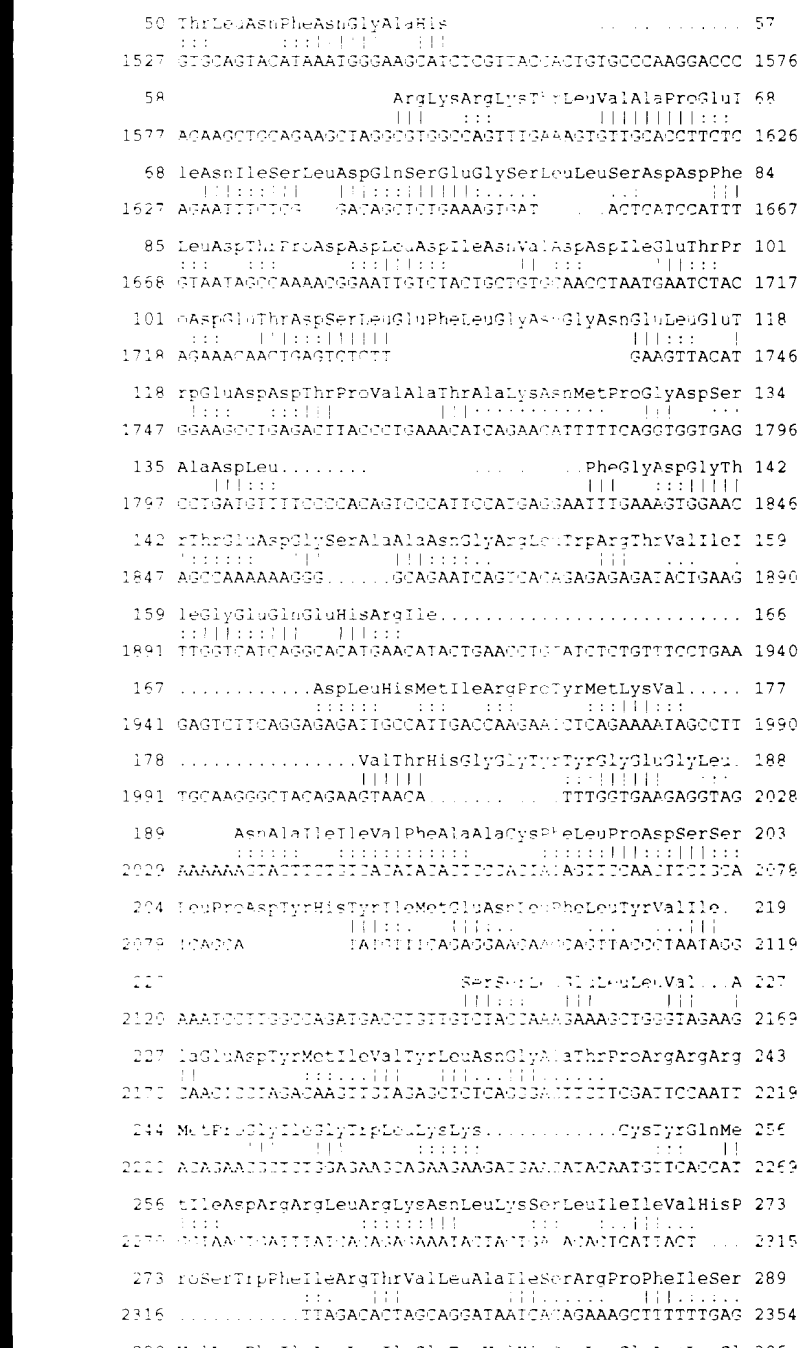
alignment_block:
 US-09-174-937-2 x Q12261   ..

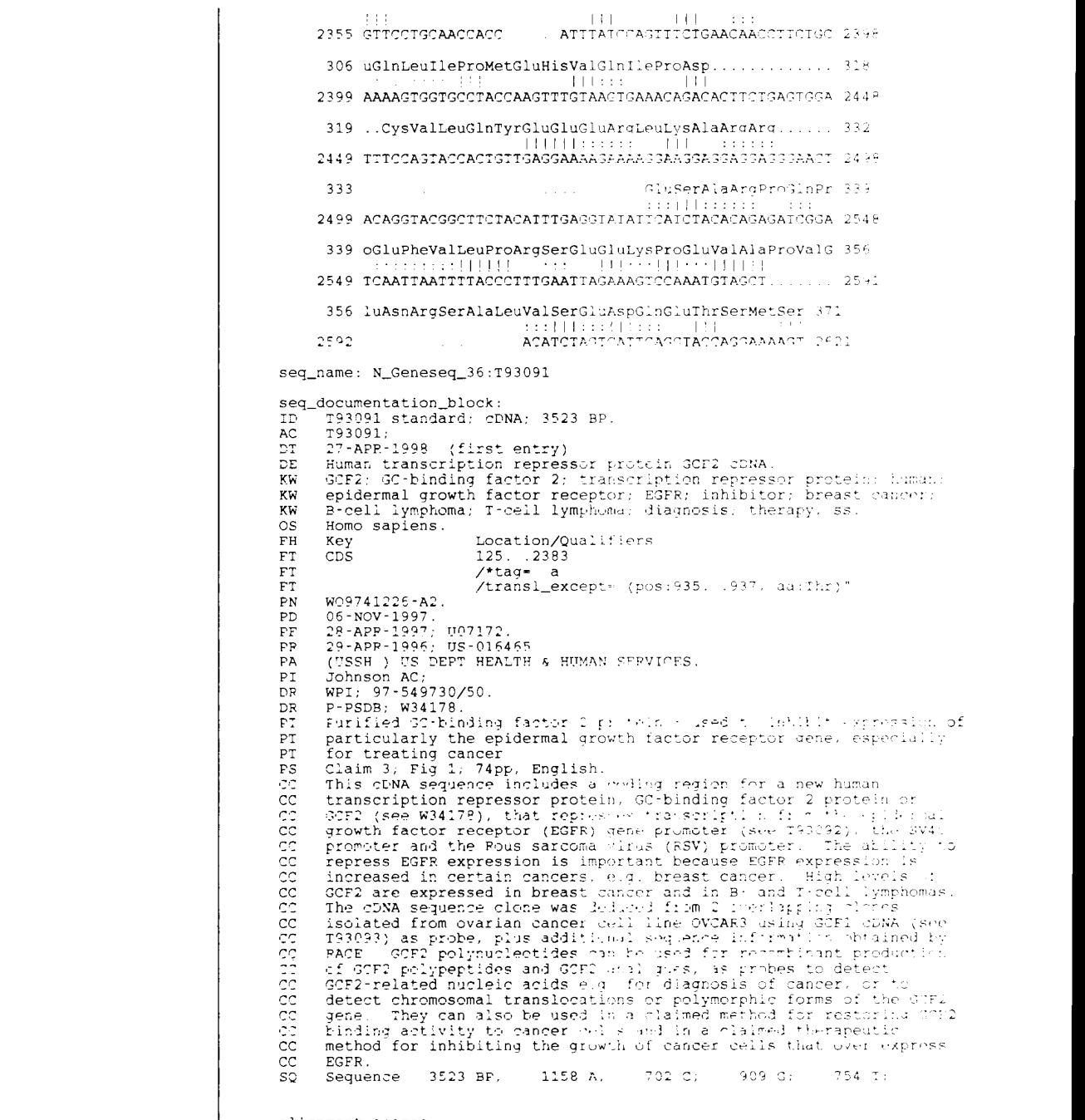
 Align seg 1/1  to: Q12261  from: 1  to: 8224

     11 GluAsnValAspValLysGluGluTrpGlnAspGluAspLeuProArgPr 27
         |||......:|||        .....:   :::       |       :::
   1380 GAAGAAACAGATCCAGTGCATGATCIAATGGTGAAATTTTACCTAA.. 1427

     27 oLeuProGluGluThrGlyValGluLeuLeuGlySerPro.......... 40
         :::|||:::           ::::::
   1428 .TTCCCTGACATAATTGAAATACACCTATACCACAGTGAACAAAATGAAG 1476

     41                ..............ValGluAspThrSerSerProProAsn 41
                                      :::      ::::::::      :::
   1477 AAGAAGAAGAAGAGTGTGCAAATGCTAGTGATGTGAGAAACCATCCCATCT 1526
```





seq_name: N_Geneseq_36:T93091

seq_documentation_block:
ID T93091 standard; cDNA; 3523 BP.
AC T93091;
DT 27-APP-1998 (first entry)
DE Human transcription repressor protein GCF2 cDNA.
KW GCF2; GC-binding factor 2; transcription repressor protein; human;
KW epidermal growth factor receptor; EGFR; inhibitor; breast cancer;
KW B-cell lymphoma; T-cell lymphoma; diagnosis, therapy, ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 125. .2383
FT /*tag= a
FT /transl_except= (pos:935. .937, aa:Thr)"
PN WO9741226-A2.
PD 06-NOV-1997.
PF 28-APP-1997; U07172.
PR 29-APP-1996; US-016465
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Johnson AC;
DR WPI; 97-549730/50.
DR P-PSDB; W34178.
PT Purified GC-binding factor 2 p; [illegible] used t. [illegible] expression of
PT particularly the epidermal growth factor receptor gene, especially
PT for treating cancer
PS Claim 3; Fig 1; 74pp, English.
CC This cDNA sequence includes a coding region for a new human
CC transcription repressor protein, GC-binding factor 2 protein or
CC GCF2 (see W34178), that represses transcription from the [illegible]
CC growth factor receptor (EGFR) gene promoter (see T93092), the SV4[illegible]
CC promoter and the Rous sarcoma virus (RSV) promoter. The ability to
CC repress EGFR expression is important because EGFR expression is
CC increased in certain cancers, e.g. breast cancer. High levels of
CC GCF2 are expressed in breast cancer and in B- and T-cell lymphomas.
CC The cDNA sequence clone was deduced from 2 overlapping clones
CC isolated from ovarian cancer cell line OVCAR3 using GCF1 cDNA (see
CC T93093) as probe, plus additional sequence information obtained by
CC RACE. GCF2 polynucleotides can be used for recombinant production
CC of GCF2 polypeptides and GCF2 fusion genes, as probes to detect
CC GCF2-related nucleic acids e.g. for diagnosis of cancer, or to
CC detect chromosomal translocations or polymorphic forms of the GCF2
CC gene. They can also be used in a claimed method for restoring GCF2
CC binding activity to cancer cells and in a claimed therapeutic
CC method for inhibiting the growth of cancer cells that over express
CC EGFR.
SQ Sequence 3523 BP; 1158 A; 702 C; 909 G; 754 T;

alignment_scores:

```
           Quality:   91.00              Length:    148
             Ratio:  1.123                Gaps:      6
 Percent Similarity:  54.730   Percent Identity: 22.297

alignment_block:
 US-09-174-937-2 x T98091     .

 Align seq 1/1  to: 198091  from: 1  to: 2523

      14 AspValLysGluGlu.....TrpGlnAspGluAspLeuProArgProLe 28
         ||||||||||:::|||     :::|||:::   :||||
    1724 GATGTTAAAAAAGAGIIAACGIATCAGAACACAGATTIA..... 1762

      28 uProGluGluThrGlyValGluLeuLeuGlySerProValGluAspThrS 45
                                  ||:   :::::::::   :
    1763 ...     .       .      .. .....AGTGAAATTAAGGAAGAAG 1781

      45 erSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLysArgLys 61
         :::::    :: :||!    :     ::::::    ::::::::::
    1782 AGCAGGTAAAGTCTACTGACAGAAAGTCAGCAGTGGAAGCCCAAAACGAG 1831

      62 ThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerGluGlySe 78
         :::         |||:::   :::|||:::   ::::::|||||||:::
    1832 GTGACTGAAAATCCAAAACAGAAAATTGCAGCAGAAAGCAGTGAAAAT.. 1879

      78 rLeuLeuSerAspAspPheLeuAspThrProAspAsp..... 90
                          ::.|||   |||:.:::
    1880 ...................GTTGATTGTCCGGAGAATCCTAAAATTAAGT 1910

      91 ...........LeuAspIleAsnValAspAspIleGluThrProAspGlu 103
                  ||||||   :::   |||| |:::::      :::
    1911 TGGATGGAAAACTTGACCAAGAAGGTGATGATGTACAACCAGCAGCTGAG 1960

     104 ThrAspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAs 120
                 |||   |||.....|||...   |||.....:|||||
    1961         .....GAGGTACTAGCTGATGGAGACACATTAGATTTTGAGGA 1998

     120 pAspThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspL 137
         |||||||      :::::::::   :::      |||:::::   |||
    1999 TGACACCGCTCAATCATCAGGCCCGAGGGCTGGTGGTGAAGAATTAGAT. 2047

     137 euPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGly 151
            ...|||:.....  |||:::      :::|||
    2048 ........GAAGGTGTTGCAAAAGATAATGCTAAAATAGATGGT 2083

seq_name: N_Geneseq_96:V52220

seq_documentation_block:
ID   V52220 standard; DNA; 18436 BP.
AC   V52220;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:87
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI; 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 693-704; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S pneumoniae
CC   genome (SEQ ID NO:1 to 391) which the [illegible] molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) [illegible] DNA [illegible] produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, in expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence   18436 BP;    6012 A;    4288 C;    4088 G;    3948 T;

alignment_scores:
           Quality:   89.50              Length:    26
             Ratio:  0.683                Gaps:      1
 Percent Similarity:  49.810   Percent Identity: 21.243

alignment_block:
 US-09-174-937-2 x V52220/rev    ..

 Align seq 1/1  to reverse of: V52220  from: 1  to: 18436

      59 ArgLysArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGl 74
         |||| |      ...|' ::::::          :::   :::
   17131 CGTAAGGACGACAATTTGACGAGTAAAGAA.....CAAACACTTAACA 17088

      74 nSerGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspL 93
         |:::|'|   |||:::     :::..         :::
   17087 AAAAGAACAAAGTATT.....TCTGATAGAGCTAAAAACCTTGATGCGC 17044

      91 euAspIleAsnValAspAspIleGluThrProAspGluThrAspSerLeu 107
         :::   ........:::::::  '          '  :::::::   '
   17043 GTGAAGAGCAATTAGAGGAAGTCGAAAGACAAAAAGAAGCAGAA...CTA 16997

     109 GluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProVa 124
         ||    .:|||    :::::   ''  ::: :   :::  :   ::
   16996 GAGCGTATTGGTGCGCTGTCTCAGGTAGAAGCAGAGATATTATCTTGGC 16947

     124 lAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuPheGlyAspG 141
         :   |||   :::|||:::     :::   |  :::   :::   ::::
   16946 TCATACAGAGGAAAACTTTACCAGCCAGTTCCAGTCCCAGTCCCAGCAG 16897

     141 lyThrThrGluAspGlySerAlaAlaAsnGlyArgGluTrpArgThrVal 157
         ::      |||
   16896 CTGAGCAAGAG ..................................... 16886

     158 IleIleGlyGluGlnGluHisArgIleAspLeuHisMetIleArgProTy 174
          :::   |||:::::   :::::::          :::::::
   16885 ...GTCAAGGAACGTTCTGACAAAATGGCCAAGGACATCTTGGTTAAGC 16839

     174 rMetLysValValThrHisGlyGlyThrThrGlyGlnGlyLeuAspAlaI 191
          ||:::   ::::::   '||     :::   :::   :::
   16838 TATGCAACGTATCGCT   CGTCAATATGTAGCGGATTCAACAAATTCAA 16792

     192 leIleValPheAlaAlaCysPheLeuProAspSerSerLeuProAspTyr 207
         :::                    |   ::::::::::
   16791 CAGTT...  .          .CATCTGCCAGACGATACTATGAAGGTACGC 16757

     208 HisTyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLeu..... 222
                 |||:::    :::  :::::::::::'''''
   16756 ATTATTGGTCGTGAAGGTCGTAACATTCGTACCTTTGAAAGTTTGACAGG 16707

     223 ....GluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAspGlyA 238
         ............:::  :::          :::
   16706 GTTGATGTGATTATCGACATACAGCCAGAACTGCTTACCTTGTCAGGAT 16657
```

```
  238 laThrPro...ArgArg.ArgMetProGlyIleGlyTrpLeuLysLysCy 253
16656 ITGAICCGAITCGICGIGAGAITGCCCGIATGAIIATGG   AAATGTTG 16610

  253 sTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSerLeuIleI 270
16609 CICAAAGATGGTC                        GTATACATCCAGC 16584

  270 leValHisProSerTrpPheIleArgThrValLeuAlaIleSerArgPro 286
16583 ICGIAICGAAGAGTTGGTTGAGAAAAACCGTTAAGAGA           . 16546

  287 PheIleSerValLysPheIleAsnLysIleGlnTyrValHisSerLeuGl 303
16545 ...TTGACAATAAGATTCGTGAATACGGTG .        ........ 16519

  303 uAspLeuGluGlnLeuIleProMetGluHisValGln 315
16518 ..........AGGCTGCTGCCTATGAAATTGGTGGC 16492
```

seq_name: N_Geneseq_36:Q29472

seq_documentation_block:

```
ID   Q29472 standard; DNA; 4118 BP.
AC   Q29472;
DT   04-MAR 1993  (first entry)
DE   Muramidase released protein gene.
KW   MRP, detection, prevention; screening; diagnostic; ss.
OS   Streptococcus suis type II pathogenic.
FH   Key             Location/Qualifiers
FT   cds             159..3929
FT                   /*tag= a
FT   signal_peptide  159..299
FT                   /*tag= b
FT   mat_peptide     300..3296
FT                   /*tag= c
FT   promoter        4..9
FT                   /*tag= d
FT                   /note= "-35 region"
FT   promoter        29..34
FT                   /*tag= e
FT                   /note= "-10 region"
FT   promoter        40..45
FT                   /*tag= f
FT                   /note= "-35 region"
FT   promoter        63..68
FT                   /*tag= g
FT                   /note= "-10 region"
FT   rbs             147..152
FT                   /*tag= h
FT                   /note= "ribosome binding site"
FT   misc_feature    4069..4080
FT                   /*tag= i
FT                   /note= "dyad symmetry region"
FT   misc_feature    4087..4098
FT                   /*tag= j
FT                   /note= "dyad symmetry region"
FT   misc_feature    2757..3014
FT                   /*tag= k
FT                   /note= "proline-rich region"
FT   misc_feature    3825..3926
FT                   /*tag= l
FT                   /note= "membrane anchor region"
FT   repeat_unit     3015..3176
FT                   /*tag= m
FT                   /note= "repetitive unit"
FT   repeat_unit     3423..3584
FT                   /*tag= n
FT                   /note= "repetitive unit"
FT   repeat_unit     3585..3743
FT                   /*tag= o
FT                   /note= "repetitive unit"
PN   WO9216630-A.
PD   01-OCT-1992.
PF   19-MAR-1992; NL0054.
PR   21-MAR-1991; NL-000510.
PA   (DIER-) CENT DIERGENEESKUNDIG INST.
PI   Smith HE, Vecht U;
DR   WPI; 92-349215/42.
DR   P-PSDB: R27746.
PT   Deoxyribonucleic acid encoding virulence characteristic of
PT   Streptococcus suis - useful for antibody and polypeptide for
PT   diagnosing and preventing infections in pigs and humans
PS   Claim 3, Fig 2, 86pp; English.
CC   The sequence is that of the muramidase released protein gene from
CC   Streptococcus suis type II (pathogenic) which allows the detection
CC   and the prevention of infections by S. suis in a more effective
CC   manner than was previously possible. It facilitates screening of
CC   e.g. pigs and elimination of infected and carrier pigs can then be
CC   carried out. The new diagnostic tests can distinguish between
CC   avirulent and virulent strains. It may be used in the prodn. of
CC   a vaccine. See also Q29470 and Q29471.
SQ   Sequence   4118 BP;    1392 A;    770 C;    878 G;    1078 T;
```

alignment_scores:

```
            Quality:   89.00          Length:    305
              Ratio:   0.659            Gaps:     10
 Percent Similarity:  44.262  Percent Identity:  18.361
```

alignment_block:

```
 US-09-174-937-2 x Q29472 ..

 Align seg 1/1  to: Q29472  from: 1  to: 4118

   22 GluAspLeuProArgProLeuProGlnGluThrGlyValGluLeuLeuGl 38
 3078 GAAGAGGGAACAAAACCAAACAAATCAATCCCAGGTTACGAGTTCACAGG 3127

   38 ySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsnG 55
 3128 TAAAACTGTTACTGACGAAGATGCC.....................AACA 3156

   55 lyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsn...... 69
 3157 CAACTCACATCTACAAGAAAACA.........CCAGAAGTTAAGAATGGT 3197

   70 ......IleSerLeuAspGlnSerGluGlySerLeuLeuSerAspAs 83
 3198 ACAGTTGTTGTTAACTATGTAACAGAAGATGGCACAGTTATCAAGGAACC 3247

   83 pPheLeuAspThrProAspAspLeuAspIleAsnValAspAspIleGluT 100
 3248 TGTAACAGATACACCAACTTCT                            3269

  100 hrProAspGluThrAspSerLeuGluPheLeuGlyAsnGlyAsnGluLeu 116
 3269 ...............................................   3269

  117 GluTrpGluAspAspThrProValAlaThrAlaLysAsnMetPro..... 131
 3270 ......CCAGAAGGCACACCATACGACACTACAGACAACAAACCTAAGAC 3313

  132 ........GlyAspSerAlaAspAlaPheGlyAspGlyThrThrG 144
 3314 AATCACTTTCAAAGGTGAAGAGTATGAATTGGTTCGTGTTTACAGTACAG 3363

  144 luAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrValIleIleGly 160
 3364 AAAACGGTAAAGTTGTAGAAGGTGAAACAGTTGTGACTTACGTTTACCGT 3413

  161 GluGlnGluHisArgIleAspLeuHisMetIleArgProTyrMetLysVa 177
 3414 AAAGTCGAA.........             ACACCTGCTAAGAAAGT 3439
```

```
     177 lValThrHisGlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleValP 194
         |||||||:::      :::      |||      |||   :::
    3440 TGTAACTAAC....  CACGTTGATGAAGAGGGTAACCCTGTTGCGCCGC 3483

     194 heAlaAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrIle 210
                       |||...:::|||   : :   |||:  .  .
    3484 AAGAAGAGGGAACAAAACCAAACAAATCAATCIIAGGTTACGAATTT .. 3530

     211 MetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAl 227
                                     :::   :::   ::::::
    3531 ......              . . . . . ACAGGTAAAACTGTTACTGA 3550

     227 aGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgM 244
         |||:||         :::::::::         ||||||      ::::
    3551 CGAAGATGGCAACACAACTCACATCTACAAGAAAACACCTGCTATGAAAG 3600

     244 ctProGly.IleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgAr 260
         ::              :::|||||          .. ...   .:::
    3601 TTGTGACTAACCACGTTGATGAAGAAGGTAACCCTATTGCTCCACAAGAG 3650

     260 gLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpPheI 277
         |:::   ::::::   :::       :::   ||:             :
    3651 GATGGGACAACACCAAAACGTCAAATTTCAGGTTACGAGTATGTGCGTAC 3700

     277 leArgThrValLeuAlaIleSerArgProPheIleSerValLysPheIle 293
         ::   :::      :::::|||  :::::   :::::|||:::
    3701 TGTAGTTGATGAAGAAGGTAACACGACACATATTTATCGCAAACTTTCTA 3750

     294 AsnLysIleGlnTyrValHisSer......LeuGluAspLeuGluGlnLe 308
         :..   |||: .  .           :|:. .:::|||:::||:::
    3751 ATAAACCAACAACACCTGAGAAGGAAACTCCTGCAAAACCTCAAGCAGGT 3800

     308 uIleProMetGlu 312
         :   |||:::::
    3801 AAAACCGCTTCAG 3813

seq_name: N_Geneseq_36:V40629

seq_documentation_block:
ID   V40629 standard; cDNA; 6318 BP.
AC   V40629;
DT   27-OCT-1998 (first entry)
DE   Musca domestica voltage-sensitive sodium channel gene.
KW   voltage-sensitive sodium channel; insecticide, sensitivity;
KW   resistance; Vssc1 gene, ss.
OS   Musca domestica.
FH   Key             Location/Qualifiers
FT   CDS             1..6318
FT                   /*tag= a
FT                   /product= voltage-sensitive sodium channel
PN   WO9828446-A1.
PD   02-JUL-1998.
PF   18-DEC-1997; U24256.
PR   24-DEC-1996; US-772512
PA   (CORR ) CORNELL RES FOUND INC.
PI   Ingles PJ, Knipple DC, Soderlund DM;
DR   WPI; 98-377674/32.
DR   P-PSDB; WS7772.
PI   New isolated voltage-sensitive sodium channel polypeptides -
PI   obtained from house flies, which are capable of conferring
PI   sensitivity or resistance to an insecticide in insects
PS   Claim 5; Page 42-45; 96pp; English.
CC   The sequence is that encoding a voltage-sensitive sodium channel
CC   (VSSC) of Musca domestica (NAIDM strain). Such a VSSC is capable of
CC   conferring sensitivity or resistance to an insecticide. It can be used
CC   to screen a cDNA library for other VSSC genes. Antibodies raised
CC   the VSSC can be used to detect VSSCs and these can be used in
CC   drug screening. Antisense nucleic acids and vectors containing
CC   the sequence may be used to reduce VSSc expression in an insect.
CC   The VSSCs can be used for conferring sensitivity or resistance to
CC   insecticides such as DDT and analogues and pyrethroids in insects
CC   such as house flies, fruit or vinegar flies, tobacco budworm,
CC   Colorado potato beetle, German cockroach or yellow fever mosquito.

SQ   Sequence   6318 BP;   1713 A;   1341 C;   1597 G;   1667 T;


alignment_scores:
            Quality:   88.50              Length:    [illegible]
              Ratio:   0.651                Gaps:    [illegible]
 Percent Similarity:   48.057   Percent Identity:   22.[illegible]

alignment_block:
 US-09-174-937-2 x V40629   ..

 Align seq 1/1  to: V40629  from: 1  to: 6318

      11 GluAsnValAspValLysGluGluTrpGlnAspGlu...AspLe.ProAr 26
         |||.....:::  :::  |||.::  :::  :::  :::        ::
    3402 GAAGAAATCAAAATTCATGAACAACACAACGATGATTGGAAATT AATAA 3451

      26 gProLeuProGluGluThrGlyVal.LeuLe.....................wAsp  8
         :|||     :::    .||||:::  :::  :  :  :::
    3452 ACCACCAAGACAATAGACTGGAACATGAGTAAACCATAGAGGTTTGTCC 3501

      39 SerProValGluAspThrSerSerProProAsnThrLeuAsnPh/AsnGl 55
               :::::::| |::::|               :::
    3502 ATACAGGACGATGACACTGGCAGC.............ATTAACTCATATGG 3549

      55 yAlaHisArgLysArgLysThrLeuValAlaProGl.IleAsnLeuSerL 72
         |...||||:::..|||
    3540 TAGCCATAAGAATCGACCATTC.............................. 3561

      72 euAspGlnSerGluGlySerLeuLeuGlyAspAspPheLeuAspThrPro 89
            ::::::'|||::.          :::  :::  :::::
    3562 ...  AAGGACGAGAGCCACAAGGGTAGGGCCAAGACATCCAGTGCGAG 3606

      89 AspAspLeuAspIleAsnValAspAspIleGluThrProAspGl.ThrAs 1 5
         ...     |||::::::   ::::  :::       :::
    3607 GAGAAACGCGACGTCAGCAAAGAGGATTTGGGTTTCGATCAGGTATTTCA 3656

     105 pSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGl........ 119
         |..:   |||      |||:::  ::::::  :::   :::
    3657 CGAG...GAGGTCGAGGGCGATGAGGTTAGTTGGATGGTGACGTTATCA 3706

     120 ..............................AspAspThrProVal....... 124
                                      :  :  :::
    3704 TTCATGCGCAAAACGACGACGAGATAATTGATGACTATTTGGGTTATTT 3753

     125 .........AlaThrAlaLysAsnMetPro.........GlyAspSerAl 139
                 :::      '  :::::::  '          :::
    3754 TTCCCCGACTCGTACTACAAGAAGTTTCCGATCTTGGCCGGCGATGAGGA 3803

     135 aAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyA 152
         :::   |||   ::::||                      ::::
    3804 CTGGCCGTTCTGGCAAGGATGG.................GGCAATTTAC 3853

     152 rgLeuTrpArgThrVal.IleLeuGlyGlyGlnLeH.sAr.GluAspLeu 168
         |||||   .:.|||       :::    : ::::::
    3836 GACTG...AAAACTTTTCAATTAATTGAAAATAAA.............. 3867

     169 HisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrTyrGl 185
                                                  ::
    3868 ..............  ..............................TATTTTGA 3875

     185 yGluGlyLeuAsnAlaIleIleValPheAla......AlaCysPheLeuP 200
         .......  :::::::|||::::::::::       :::::
    3876 AACCCCAGTTATCACTATGATTTTAATCAGTACTTAGCTTTGGCCTTAG 3925

     200 roAspSerSerLeuPro.................AspTyrHisTyrIle 216
            |||      ||||'|'               '  '   '
    3926 AAGATGTTCATTTACCCGATCGATCTGTCATGCAGGATATACTGTACTAC 3975

     211 MetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAl 227
         |'|.....:::|||      |'''  '       '  :::  :::
    3976 ATGGACAGGATATTTACG...GTGATATTCTTTTTGTAGATGTTGATC... 4021
```

```
 227 aGluAspTyrMetIleVal.TyrLeuAsnGlyAlaThrProArgArgArg 243
          :::::: ::: ||':::  ::::::|||
4021     AAA|GG|GGCCTGGGCTTTAAGGTTTA'TTCACCA        4057

 244 MetProGlyIleGlyTrpLeuLys      LysCysTyrGlnMet 256
     |||||||||:::|||||| .           :|||| ..
4058 ATGCCTGGTATTGGCTGGATTTCGTGATTGTCATGCTATCGCTTA 4102
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 07:04:08 ; Search time 5888.89 Seconds
(without alignments)
-573.876 Million cell updates/sec

Title: US-09-174-937-3
Perfect score: 1113
Sequence: 1 atggggaccaccgaagccac.....atcaggaaacaagcatgtcc 1113

Scoring table: IDENTITY_NUC

Searched: 921193 seqs, 1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold . 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	318.2	28.6	5654	9	AB002365	AB002365 Human mRN
2	268.8	24.2	2223	12	AF038207	AF038207 Mus muscu
3	231.6	20.8	1396	5	AR028453	AR028453 Sequence
4	231.6	20.8	2992	11	HSU58178	U58178 Homo sapien
c 5	222	19.9	179301	41	AC011498	AC011498 Homo sapi
6	102.8	9.2	1430	9	HSRHOGAPX	Z29024 H.sapiens r
7	102.8	9.2	3347	9	HSU58878	U58878 Human CDC42
8	99.6	8.9	169798	44	AC009556	AC009556 Homo sapi
9	99.8	8.1	126121	43	AC011469	AC011469 Drosophil
c 10	83.6	7.5	169798	44	AC009556	AC009556 Homo sapi
c 11	80.4	7.2	78170	43	AC015994	AC015994 Drosophil
12	76.4	6.9	179301	41	AC011498	AC011498 Homo sapi
13	68.2	6.1	1378	10	HSU58878	U58878 Human CDC42
c 14	54.4	4.9	6573	1	AB027763	AB027763 Streptomy
15	52	4.7	3624	5	AR028609	AR028609 Sequence
16	52	4.7	3624	5	I41422	I41422 Sequence 6
c 17	49.8	4.5	10112	2	AE002046	AE002046 Deinococc
c 18	47.6	4.3	40909	1	SC135	AL031541 Streptomy
c 19	47	4.2	3281	24	LMA241459	AC241459 Leishmani
c 20	46.8	4.2	44378	1	SCC9	AL079856 Streptomy
21	46.8	4.2	4773	40	HS541725	Z41725 Human prote
22	46.6	4.2	5396	9	HSPLM9	Z78679 H.sapiens g
23	46.6	4.2	2690	9	HSPLKPHL	Z34974 H.sapiens m
24	46.8	4.1	31360	1	SC90C7	AL035161 Streptomy
25	45.8	4.1	2225	12	AB006016	AB006016 Mus muscu
c 26	45.8	4.1	428	12	BB7552	BB7552 BB7552:57939
c 27	45.6	4.1	5973	1	AB024056	AB024056 Streptomy
28	45	4.0	2618	9	HSB6PR	X79249 H.sapiens b
29	44	4.0	8200	1	STMPKCYS	U17 66 Streptomyce
c 30	43.8	3.9	22115	1	SCC22	AL049839 Streptomy
31	43.8	3.9	80095	35	AC004283	AC004283 Drosophil
c 32	43.8	3.9	162666	41	AC007701	AC007701 Drosophil
33	43.6	3.9	38532	1	SCE94	AL049626 Streptomy
34	43.6	3.9	3044	5	AR016557	AR016557 Sequence
35	43.6	3.9	3044	5	AR021277	AR021277 Sequence
36	43.6	3.9	3044	5	AR036871	AR036871 Sequence
37	43.6	3.9	3044	5	I30447	I30447 Sequence 44
38	43.6	3.9	3044	5	I35691	I35691 Sequence 44
39	43.6	3.9	4240	11	HSU67739	U67739 Human cGMP-
40	43.4	3.9	10918	2	AE002043	AE002043 Deinococc
41	43.2	3.9	2601	9	PTPLAKPH	Z37975 B.taurus mR
c 42	43.2	3.9	169604	33	AC007186	AC007186 Drosophil
c 43	43	3.9	40740	1	SC10J	AL109848 Streptomy
44	43	3.9	2449	1	STMARC8	U19401 S. lavendul
45	43	3.9	5047	9	AB011161	AB011161 Homo sapi

ALIGNMENTS

RESULT 1
AB002365
LOCUS AB002365 5654 bp mRNA PRI 14-FEB-1999
DEFINITION Human mRNA for KIAA0367 gene, partial cds.
ACCESSION AB002365
VERSION AB002365.1 GI:2224674
KEYWORDS KIAA0367.
SOURCE Homo sapiens male brain cDNA to mRNA, clone:HH0141.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:dnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349994
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..>2151
                     /gene="KIAA0367"
     CDS             <1..>2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20922.1"
                     /db_xref="GI:2224675"
                     /translation="ALFDGDPHLSTENPALVPDALLASDTCLDISEAAFDHSFSDASG
                     LNTSTGTIDDMSKLTLSEGHPETPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDLL
                     RFPEHFLYGGDPPLEEDSLKQSLAPYTPPFDLSYLTEPAQSAETIEEAGSPEDESLGC
                     PAAEIVLSAIPDRPSEGNQAFTKNRLPGSQLAVLHIREDPESVYLPVGAGSNILSPSN
                     VDWEVETDNSDLPAGGDIGPPNGASKEIPELEEEKTIPTKEPEQIKSEYKEERCTEKN
                     EDRHALHMDYILVNREENSHSKPETCEERESIAELELYVGSKETGLQGTQLASFPDTC
                     QPASLNERKGLSAEKMSSKGDTRSSFESPAQDQSWMFLGHSEVGDPSLDAPDSGPGWS
                     GKTVEPFSELGLGEGPQLQILEMKPLESLALEEASGPVSQSQKSKSPGPAGPDAVTH
                     DSEWEMLSPQPVQKNMISDTEMEEETEFLELGTRISRPNGLLSEDVGMDIPFEEGVLS
                     PSAADMPPEPPNSLDLNDTHPRFIKLTAPNINLSLDQSEGSILSDDNLDSPDEIDINV
                     DFLDTPDEADSFEYTGHDPTANKDSGQESFSIPEYTAEEEPEDNPLWPTVVIGEQEQP
                     IDMKVIEPYPPVISHGGYVGPGLNAIIVFAACFLPDSSRADYHYVMENLFLYVISTLE
                     LMVAEDYMIVYLNGATPKKEDAPAPLDEEMIPDD"
BASE COUNT     1664 a   1188 c   1198 g   1604 t
ORIGIN


  Query Match             28.6%;  Score 318.2;  DB 9;  Length 5654;
  Best Local Similarity   69.2%;  Pred. No. 3.1e-54;
  Matches  559;  Conservative    0;  Mismatches  288;  Indels   37;  Gaps    3;

Qy      97 ggggtggaactgcttggcagccccggtggaagacacatcctctcctcccaacacggctaaat 156
           |  ||  |  |||  |  |||  | |  ||     |   |     ||||| ||  | ||  ||
Db    1510 GAGGGGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCTCCTAATTCTCTGGAT 1569

Qy     157 ttcaacggagcgcatcgtaagaggaagacgctggtggccccagagatcaacatttctctg 216
           | || |   | |||| |  ||| |  | |||    |||||| | |||||  ||||||||
Db    1570 CTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTG 1629

Qy     217 gatcagagtgagggtcctgtgtccgatgacttcttggatacccctgatgacctggat 276
           || || ||||| || ||  | || || |||||   ||||| |  || |||||  | ||
Db    1630 GACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGAC 1689

Qy     277 attaacgtggatgacatcgaggaccccgatgaagaccgactcgctggagttcctgggaat 336
           || || ||||||||  | || ||||||||||||  | || ||  | |||| |   ||
Db    1690 ATCAATGTGGATGAACTTGATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGG---- 1745

Qy     337 ggcaacgaactggagtgggaagacgacaccccgtggccaccgccaagaacatgcccggg 396
                                        ||  |  |||| ||||| ||     | ||
Db    1746 ----------------------------CCATGATCCCACAGCCAACAAAGATTCTGGC 1776

Qy     397 gacagcgcggatctatttggggacggcacgacggaggacggcagcgccccaaaaaaac 456
           |     | ||     ||   ||    ||| | || || |   | |  | |||  |
Db    1777 CAAGAGTCAGAGTCTATTCCAGAATATACGGCCGAAGAGGAACGGGAGGACAAC---CGC 1833

Qy     457 ctgtggcggacagtgatcatggtgatcaatgatgcacgtatgaaatgtgatgtgg 516
           || ||| |||||||| |||| || ||||||||| || || ||| || |  |||
Db    1834 CTTTGGAGGACAGTGGTCATTGCAGAACAAGACCACCCATTCACATCAACTCCATCCAC 1893

Qy     517 ccttacatgaaagtggtcacccacaggagggtactacgggcgaaggcctcaacgccatcatc 576
           || |||| || |||  |  | |||||||| |||| || || || || || ||||||
Db    1894 CCCTACAGGAGAGTCATTTCTCACGGAGGATACTATGGGGACGGTCTAAATGCCATCATT 1953

Qy     577 gtcttcgcagcctgcttccttccagacagcagcctccccgactaccactacatcatggag 636
           || || || ||||| || || |||||||||||| |   | || ||||||||  ||||||
Db    1954 GTGTTTGCCGCCTGTTTTCTGCCAGACAGCAGTCGGGCGGATTACCACTATGTCATGGAA 2013

Qy     637 aacctcttcctgtacgtcatcagcagcttagagctcctggtggctgaggactacatgatc 696
           || || ||||| || || || ||  | ||||| |  |||| |||| |||| ||||
Db    2014 AATCTTTTCCTATATGTAATAAGTACTTTAGACTTGATGGTAGCTGAAGACTATATCATT 2073

Qy     697 gtgtacctgaacggtgccacgccc-aggcggagatgcctggaatggatctgctgaggaa 755
           |||||| |||| ||||| || |||  |  |||||||||| ||  | |||||| ||||||
Db    2074 GTGTACTTGAATGGTGCAACCCCCAAGAAGGAGGATGCCAGGGCTACCCTGGATGAAGAA 2133

Qy     756 gtgctaccagatgatcgacggaggttgcggaaaaacctgaagtccttgatcatcgtcca 815
           ||||||||||||||| ||| |  |||| |||| ||  ||| || || |||| | ||
Db    2134 ATGCTACCAGATGATTGACAGACGGTTGAGGAAGAATTTGAAATCATTCATCATTGTTCA 2193

Qy     816 ccctcgtggttcattcggactgtgctggccatctctgcgcctttcatcagcgtcaagtt 875
           || || |||||||  | ||  | || ||  |  | || |||| || ||    ||  |
Db    2194 TCCATCTTGGTTCATCAGAACAATCCTTGCTGTGACAGCACTTTTATAACTTCAAAATT 2253

Qy     876 catcaacaagatccagtacgtgcaagagcttggaagacctggagcaactcatccctatgga 935
           || ||  || ||  | || ||  | |||||  ||| ||      || ||||| ||||
Db    2254 CAGCAGTAAAATTAAATATGTCAAAAGCTTATCAGAACTCAGTGGGCTGATCCCAATCAA 2313

Qy     936 acacgtccagatcccagactgcgtcctgcaatacgaagaggaa 978
             | |||| || |||||  || || |  ||   || || |||
Db    2314 TTGCATCCACATTCCAGAGAGCATCATCAAACTGGATGAAGAA 2356


RESULT   2
AF035207
LOCUS       AF035207     2213 bp    mRNA            ROD       25-FEB-1998
DEFINITION  Mus musculus Nip2l mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip2l gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, CuraGen Corporation,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1..2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit59 and D9Mit1"
     gene            1..2213
                     /gene="Nip2l"
     CDS             134..1114
                     /gene="Nip2l"
                     /note="associates with motifs of both the adenovirus E1B
```

```
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP21"
                     /protein_id="AAC04329.1"
                     /db_xref="GI:2911348"
                     /translation="MEGVELKEEWQDEDFPIPLPEDDSIEADTLDGTDPDRQPGSLEV
                     NGNFVPKFLMAPDISLTLPETEGSLWSDDLDEAGEVDLELDTFSENSDEFEWEDLP
                     KPKTIEVIPKGSITEYTATFFKGDGPPWPMFPIGEQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESGQPNYRYLMDNLFKYVIGTLELLVAENYMIIYLNGAIPP
                     KMPSLGWLPRCYQQIDPRLPKNLKSLIIVHPSWFIRTLLAVTRPFISSKFSQKIRYVF
                     NLAELAELVPMEYVGIPECIKQYEEEKFKKRQKPVDQELNGKQEPPKSEQ"
BASE COUNT      637 a    400 c    571 g    605 t
ORIGIN


  Query Match             24.2%;  Score 298.8;  DB 12;  Length 2213;
  Best Local Similarity   58.4%;  Pred. No. 2.3e-44;
  Matches  571;  Conservative    0;  Mismatches  362;  Indels   45;  Gaps    4;

Qy      9 cacgaaggccacgctccggatggaaaacgtggacgtgaaggaggaatggcaggacgaaga 68
          ||  ||  | | |||  ||||||||   |||||  |||||||||||||||| |||||
Db    115 CATTGAGTCTAGGCTGAGGATGGAAGGTGTGAGCTGAAGGAAGAATGGCAGGATGAAGA 174

Qy     69 tcttcccaggccactcccagaaagagacgggggtggaactgcttggcagcccggtggaaga 128
          | |||| |  ||  | ||||||||     | ||||     |         |      ||
Db    175 TTTTCCAATACCTTTACCAGAAGATGACAGCATTGAAGCAGATACACTAGATGGAACTGA 234

Qy    129 cacatcctctcctcccaacacgctaaatttcaacggagcgcatcgtaagaggaagacgct 188
          ||  |   |  ||   | |  || |  | ||  |  | |           |||||||  ||
Db    235 TCCAGACAGACAGCCTGGTTCTTAGAAGTTAATGG---GAACAAAGTAAGGAAGAAACT 291

Qy    189 ggtggccccagagatcaacatttctctggatcagagtgagggtccctgctgtccgatga 248
          | ||||||||| |||| | |  | |||| ||   |||| || || |||  ||||||||
Db    292 GATGGCCCCAGACATCAGCCTGACCCTGGATCTGGTGAAGGTTCTCTGTGGTCCGATGA 351

Qy    249 cttcttggataccctgatgacctggatattaacgtggatgacatcgagacccccgatga 308
          ||           ||| |    ||  | || |||| | | | || |||||    ||
Db    352 TTT---------GGATGAAGCTGGAGAGGTTGACCTGGAAGGCTTAGACACCCCATCAGA 402

Qy    309 gaccgactcgctggaagttcctggggaatggcaacgaactggagtgggaagacgacacccc 368
                              |||   | ||  | ||||||||||| ||    ||
Db    403 ----------------------GAACAGTGATGAGTTTGAGTGGGAAGATGATCTTCC 438

Qy    369 cgtggccaccgccaagaacatgcccgggacagcgcggatctatttggggacggcacgac 428
          |   |||  | |  |    ||  | || ||  |      |  |||
Db    439 CAAACCCAAAACTACTGAAGTCATTAGGAAAGGCTGGATTACTGAGTACACAGCCACAGA 498

Qy    429 ggaggacggcagccgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaaga 488
          ||  | ||           | || ||||  ||| |   | |  || || ||||| ||
Db    499 AGAAAAGGGGG---------ATGGACGCTGGCTGGCGCATGTTCAGGATTGGAGAGCAGGA 549

Qy    489 gcaccgtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggta 548
          ||| |  | ||| || |   |||| ||  |  || |||||  | | ||| ||||| ||
Db    550 CCACAGGGTGGACATGAAGGCGATCGAGTCCTACAAAAAAGTTATTAGCCATGGAGGATA 609

Qy    549 ctacggcgaaggcctcaacgccatcatcgtcttcgcagcctgcttccttccagacagcag 608
          || || || ||| | || ||||  | |  |  || | ||| ||| | || || ||  |
Db    610 TTATGGGGACGGCTTAAAIGCCATTGTTGTATTGCTGTCTGTTTCATGCCTGAGAGTGG 669

Qy    609 cctccccgactaccactacatcatggagaacctcttcctgtacgtcatcagcagcttaga 668
          |  ||  ||||    ||  | |||||  |  ||||    || || ||  |||  |||||
Db    670 TCAGCCTAACTATAGATATTTGATGGACAATCCTCTTTAAATATGTTATTGGCACTTTAGA 729

Qy    669 gctcctggatggctgaggactacatgatcgtgtcgtgaacggtgccacgcccggcggag 728
          |||  |||| || ||| |||||||||  || | ||| ||||  |  || | ||  ||
Db    730 GCTGTTGGTAGCAGAGAACTACATGATCATTTACTTAAATGGTGCGACAACTCGAAGAAA 789

Qy    729 gatgcctggaatcggctggctgaagaagtgctaccagatgatcgaccggaggttgcggaa 788
          |||||  | || |||| | ||  |||||||  || ||  |  | || ||||  ||||
Db    790 AATGCCCAGTTTGGGATGGCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGGAA 849

Qy    789 aaacctgaagtccttgatcatcgtccaccccctcgtggttcattcggactgtgctggccat 848
          ||| || | |||  | ||||| || |||||| | |||||  || || ||  | |||||  |
```

```
Db    850 AAACCTTAAGTCTCTAATCATTGTGGACCCCCTTTGG.TTA.T GAAGACTTTTGGCTAT 909

Qy    849 ctctggccctttcatcaggctcaagttctatcttcaagaatttcaatatatcaatttgga 908
          | ||| || || ||||  |  |  |  |  |  |  |  |  |  |  |  |  |||||
Db    910 TACATGACCATTTATCAGTTCAAAGTTCAGTCAGAAGAATCAGA.ACTTTTTTAACTTGAT 969

Qy    909 agacctggagtaactcatttttatttaccctttgtt.....tatttttt 948
          ||  |||| | ||  | ||
Db    970 GCACTTGCCAGAGCTTGTTCCCATTCCACTTTG.CCTCCTTTCCTTTTTTTTTTTTA 1029

Qy    949 cgaagaggaaagactgaa 966
          |||| |||  | ||
Db   1030 TGAAGAAGAAAAGTTTAA 1047


RESULT   3
AR028453
LOCUS       AR028453     1386 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 22 from patent US 5858678.
ACCESSION   AR028453
VERSION     AR028453.1  GI:5940426
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G.
  TITLE     Apoptosis-regulating proteins
  JOURNAL   Patent: US 5858678-A 22 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1..1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    336 g    373 t
ORIGIN


  Query Match             20.8%;  Score 256.6;  DB 5;  Length 1386;
  Best Local Similarity   56.9%;  Pred. No. 6.4e-37;
  Matches  551;  Conservative    0;  Mismatches  374;  Indels   45;  Gaps    5;

Qy      9 cacgaaggccacgctccggatggaaaacgtggacgtgaaggaggaatggcaggacgaaga 68
          ||  ||  | | |||  ||||||||   |||||  |||||||||||||||| |||||
Db    195 CATTGAATCCAGGCTGAGGATGGAAGGTGTGAACTTAAAGAAGAAATGGCAAGATGAAGA 254

Qy     69 tcttcccaggccactcccagaaagagacgggggtggaactgcttggcagcccggtggaaga 128
          | |||| |  ||  | ||||||||     | ||||     |         |      ||
Db    255 TTTTCCGATACCTTTACCAGAAGATGACAGCATTGAAGCAGATACACTAGATGGAACTGA 314

Qy    129 cacatcctctcctcccaacacgctaaatttcaacggagcgcatcgtaagaggaagacgct 188
          ||  |   |  ||   | |  || |  | ||  |  | |           |||||||  ||
Db    315 ACCAGAGGACCAGCCTGGCTCACTAGAAGTTAATGGAA---GAAAAGTGAGAAAGAAACT 371

Qy    189 ggtggccccagagatcaacatttctctggatcagagtgagggtccctgctgtccgatga 248
          | ||||||||| |||| | |  | |||| ||   |||| || || |||  ||||||||
Db    372 AATGGCTCCAGACATTAGCCTGACACTGGATCTGATTGATGGTCTGTATTGTCCAGATGA 431

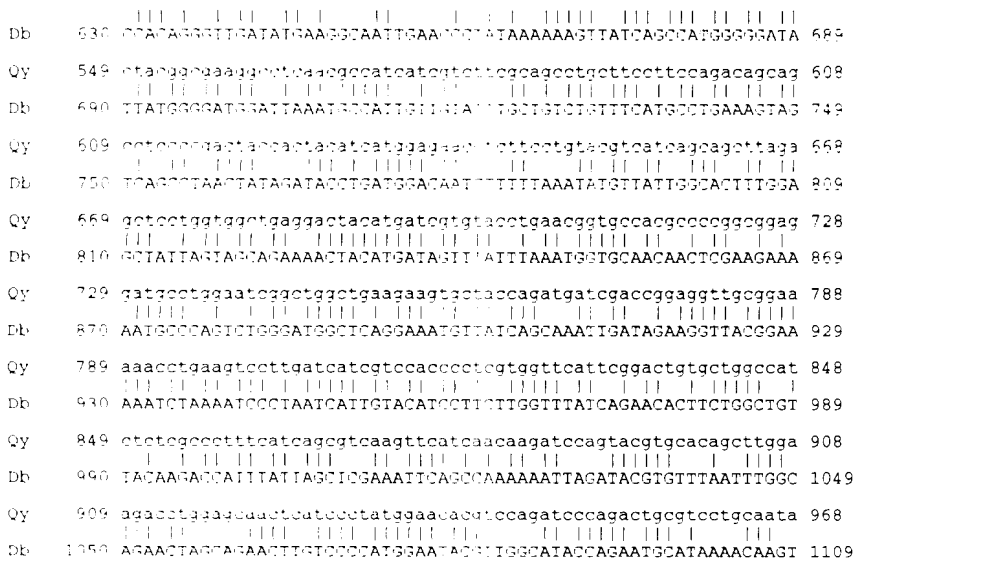
Qy    249 cttcttggataccctgatgacctggatattaacgtggatgacatcgagacccccgatga 308
          ||           ||| |    ||  | || |||| | | | || |||||    ||
Db    432 TTT---------GGATGAAAGTGGGGAGATTGACCTAGATGGCCTAGACACACCCCTCAGA 482

Qy    309 gaccgactcgctggaagttcctggggaatggcaacgaactggagtgggaagacgacacccc 368
                              |||   | ||  | ||||||||||| ||    ||
Db    483 ----------------------GAATAGTAATGAGTTTGAGTGGGAAGATGATCTTCC 518

Qy    369 cgtggccaccgccaagaacatgcccgggacagcgcggatctatttggggacggcacgac 428
          |   |||  | |  |    ||  | || ||  |      |  |||
Db    519 AAAACCCAAGACTACTGAAGTAATTAGGAAAGGCTG------AATTACTTAATACACAGC 572

Qy    429 ggaggacggcagccgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaaga 488
          ||  | ||           | || ||||  ||| |   | |  || || ||||| ||
Db    573 AGCAGAGGAAAAAGAAG---ATGGACGACTGCTGGCGCATATTTCAGGATTGGAGAACAGGA 629

Qy    489 gcaccgtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggta 548
```



RESULT 4
HSU15173
LOCUS HSU15173 2382 bp mRNA PRI 02-FEB-1998
DEFINITION Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2) mRNA, complete cds.
ACCESSION U15173
VERSION U15173.1 GI:558843
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 212 to 1156)
AUTHORS Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K., Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.
TITLE Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set of cellular proteins
JOURNAL Cell 79, 341-351 (1994)
MEDLINE 95042730
REFERENCE 2 (bases 1 to 2382)
AUTHORS Boyd,J.M.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1994) Janice M. Boyd, Institute for Molecular Virology, St. Louis University Health Sciences Center, 3681 Park Avenue, St. Louis, MO 63110, USA
FEATURES Location/Qualifiers
source 1..2382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="B-cell, S.Elledge; B-cell, J.Ambrus; fetal brain, M.Green"
/cell_type="B cells; brain"
/dev_stage="fetus (brain)"
gene 1..2382
/gene="BNIP2"
CDS 212..1156
/gene="BNIP2"
/codon_start=1
/product="BCL2/adenovirus E1B 19kD-interacting protein 2"
/protein_id="AAC00021.1"
/db_xref="GI:558844"
/translation="MEGVELLKEEWVDEDFPIPLPEDDSRAAIIAIIRRDQPGSLEV
NGNKVRKKLMAPDISLTLDPSDGSVLSDDLDESCEIDLELLTPNENSNFFEWEDDLP
KPKTTEVIPKGSITEYTAAEEKEDGPRWPMFRIGEQDHPVCMKAIEEYRKVISHGGYY
GDGLNAIVVFAVCPMPESSQPNYPYLMENLFKYVITLEIIAAMMLILNGATTRP
KMPSLGWLRKCYQQIDRRLRKNLKSLIIVHPSWFIPTLLAMVKPFISEKSQKIRYVF
NLAELAELVPMEVVSIPECIKQVDQELNGKQDEPKNDQ"
BASE COUNT 734 a 393 c 511 g 744 t
ORIGIN



Query Match 20.8%; Score 231.6; DB 11; Length 2382;
Best Local Similarity 56.8%; Pred. No. 6.2e-37;
Matches 551; Conservative 0; Mismatches 374; Indels 45; Gaps 5;

```
Qy     849 ctctcgccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttgga 908
           |  | || || || ||| || |||| |  || ||   ||||||  |  ||||
Db     988 TACAAGACCATTTATTAGCTCGAAATTCAGCCAAAAAATTACATACGTGTTTAATTTGGC 1047

Qy     909 agacctggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgcaata 968
           ||| || |  |||| |||| ||||| |  |  || ||||| ||| |    |||
Db    1048 AGAACTAGCAGAACTTGTCCCCATGGAATACGTTGGCATACCAGAATGCATAAAACAAGT 1107

Qy     969 cgaaagaggaa 979
           ||  | |||
Db    1108 TGATCAAGAA 1117


RESULT   5
AC011488/c
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *              gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *              gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *              gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *              gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *              gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *              gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *              gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *              gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *              gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *              gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *              gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *              gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *              gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *              gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *              gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *              gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *              gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *              gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *              gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *              gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *              gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *              gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *              gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *              gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *              gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *              gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *              gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *              gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *              gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *              gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *              gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *              gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *              gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *              gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *              gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *              gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *              gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *              gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *              gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *              gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *              gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *              gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *              gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *              gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *              gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *              gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *              gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *              gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *              gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *              gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *              gap of unknown length
```

```
            *    55088    81867: contig of 26780 bp in length
            *           gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *           gap of unknown length
            *   112601   179301: contig of 66701 bp in length
FEATURES             Location/Qualifiers
     source          1. .179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42598 t     85 others
ORIGIN
```

```
  Query Match             19.9%;  Score 222;  DB 41;  Length 179301;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-35;
  Matches  222;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     137 ctcctcccaacacgctaaatttcaacggagcgcatcgtaagaggaagacgctggtggccc 196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124303 CTCCTCCCAACACGCTAAATTTCAACGGAGCGCATCGTAAGAGGAAGACGCTGGTGGCCC 124244

Qy     197 cagagatcaaacatttctctggatcagagtgagggtccctgctgtccgatgacttcttgg 256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124243 CAGAGATCAACATTTCTCTGGATCAGAGTGAGGGTCCCTGCTGTCCGATGACTTCTTGG 124184

Qy     257 ataccctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccgact 316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124183 ATACCCCTGATGACCTGGATATTAACGTGGATGACATCGAGACCCCGATGAGACCGACT 124124

Qy     317 cgctggagttcctgggaaatggcaacgaactgtagtgggaag 358
           |||||||||||||||||||||||||||||||||||||||||||
Db  124123 CGCTGGAGTTCCTGGGAAATGGCAACGAACTGTAGTGGGAAG 124082
```

```
RESULT   6
HSRHOGAPX
LOCUS       HSRHOGAPX    1430 bp    mRNA            PRI       25-AUG-1997
DEFINITION  H.sapiens rhoGAP protein.
ACCESSION   Z23024
VERSION     Z23024.1  GI:312211
KEYWORDS    rhoGAP protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   Hall,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1993) Alan Hall Dr., Cell and Molecular Biology,
            Institute of Cancer, Research, Chester Beatty Laboratories, 239
            Fulham Road, London, SW3 6JB, England
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="Male"
                     /tissue_type="fibrosarcoma"
                     /cell_line="HT1080"
     5'UTR           1. .98
     CDS             99. .1418
                     /codon_start=1
                     /product="rhoGAP protein"
                     /protein_id="CAA80560.1"
                     /db_xref="GI:312212"
                     /db_xref="SWISS-PROT:Q07960"
                     /translation="MDPLSELQDDLILDDTSEALNQLKLASIDEKNWFSEEMPDFFKS
                     DDSKSSSPELVTHLKWDDPYYYDIARHQIVEVAGDDKYGPKIVFSACRMPPSHQLDHS
                     KLLGYLKHTLDQYVESDYILLYLHHGLISDNKPSLSWLEDAIEELPASLKNIKALYI
                     VHPTMFIKTLLILFKPLISPKFGQKIFYVNYLSELSPPVKLPQIPPQVLKYDDFLK
                     STQKSPATAPKPMPPPRPPLPNQQFGVSLQHLQEKNPEQEPIPIVLRETVAYLQAHALT
                     TEGIFRRSANTQVVAFVQQKYNMGLPVLPLQYNELHLPAVLLKTFLRELPEPLLTFDL
                     YPHVVGFLNIDESQRVPATLQVLQTLPEENYQVLRFLTAFIVQISAHSDQNKMTNTNL
                     AVVFGPNLLWAKDAAITLKAINPINTFIKFLLDHQGELFPSPDPSGL"
BASE COUNT      332 a    453 c    365 g    280 t
ORIGIN
```

```
  Query Match             9.2%;  Score 102.8;  DB 9;  Length 1430;
  Best Local Similarity   54.9%;  Pred. No. 3.2e-11;
  Matches  203;  Conservative    0;  Mismatches  167;  Indels    1;  Gaps    0;

Qy     605 gcagcctccccgactaccactacatcatggaggaacctcttcctgtacgtcatcaggaggct 664
           | | | |||| | ||||  | |     || ||| |  |||| | |    | |
Db     373 GAATGCCCCCCAGCCACCAGCTCGACCACAGCAAGCTCCTGGGGTACCTGAAGCACACTT 432

Qy     665 tagagctcctggtggctgaggactacatgatcgtgtacctgaacggtgccacgccccgga 724
           | ||  |    ||||      |||||||  |  |||| ||| ||  | | |  | || |
Db     433 TGGACCAGTACGTGGAGAGTGACTACACACTTCTGTATCTGCACCACGGCCTGACCAGCG 492

Qy     725 ggaggatgcctggaatcggctggctgaagaagtgctaccagatgatcgaccccgaggttgc 784
           |  | |||     || |||||||     |  |||| |  | | ||||| | ||
Db     493 ACAACAAGCCCTCCCTCAGCTGGCTCCGTGATGCCTACCGGGAGTTTGACCGCAAGTACA 552

Qy     785 ggaaaaacctgaagtccttgatcatcgtccaccccctcgtggttcattcggactgtgctgg 844
           ||| ||| | ||| |||||  |||||| || ||  |  |||||||    ||| ||| |
Db     553 AGAAGAACATCAAGGCCTTGTACATCGTGCATCCAACCATGTTCATCAAAACTCTGCTCA 612

Qy     845 ccatctctcgccctttcatcagcgtcaagttcatcaacaagatccagtaccgtgcacaggct 904
           | |||     ||  |||||||| ||||||||     | ||||||   || ||| |  |
Db     613 TCCTCTTCAAGCCCCTCATCAGCTTCAAGTTCGGGCAGAAGATCTTCTATGTGAATTACC 672

Qy     905 tggaagacctggagcaactcatccctatggaacagcgtccagatccagagctgcgtcctgc 964
           ||   || |||   | | | |    |||| ||  |  ||||||   |  | | |
Db     673 TGAGCGAGCTGAGCGAGCACGTGAAGCTGGAGCAGCTGGGGATCCCTCGCCAAGTGCTCA 732

Qy     965 aatacgaaga 974
           |||| || ||
Db     733 AATATGACGA 742
```

```
RESULT   7
HSU02570
LOCUS       HSU02570     3345 bp    mRNA            PRI       25-JAN-1994
DEFINITION  Human CDC42 GTPase-activating protein mRNA, partial cds.
ACCESSION   U02570
VERSION     U02570.1  GI:450780
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3345)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 268, 26059-26062 (1993)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 3345)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P., Self,A.J., Brill,S., van Erp,H.E.
            and Hall,A.
  TITLE     Characterization of rhoGAP: A GTPase Activating Protein for
            Rho-Related Small GTPases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3345)
  AUTHORS   Barfod,E.T.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (18-OCT-1993) Elisabeth T. Barfod, Genetech, Inc., 460
            Point San Bruno Blvd., South San Francisco, CA 94080, USA
COMMENT     On Jan 27, 1994 this sequence version replaced gi:409026.
FEATURES             Location/Qualifiers
     source          1. .3345
                     /organism="Homo sapiens"
                     /db_xref="taxon 9606"
                     /cell_line="CMK"
     CDS             <1  1391
                     /codon_start=3
                     /function="GTPase activating protein for the rho family of
                     small GTP-binding proteins that is most effective against
                     CDC42"
                     /product="CDC42 GTPase-activating protein"
                     /protein_id="AAA16142.1"
                     /db_xref="GI:409027"
                     /translation="GSSARARFLLPFCPAPLSGTGLAMDPLSELQDDLTLDDTSEALN
                     QLKLASIDEKNWPSDEMPDFPKSDDSKSSSPELVTHLKWDDPYYDIAPHQIVEVASDD
                     KYGPKIIVFSACRMPPSHQLDHSKLLGYLKHTLDQYVESDYTLLYLHHGLTSDNKPSL
                     SWLPDAYPEFDPKYKKNIKAIVIVHPTMFIKTLLILFKPLISPKFGQKIFYVNYLSEL
                     SEHVKLEQLGIPPQVLKYDDFLKSTQKSPATAPKPMPPRPPLPNQQFGVSLQHLQEKN
                     PEQEPIPIVLPETVAYLQAHALITEGIFPPSANTQVVPEVQQKYNMGLPVDFDQYNEL
                     HLPAVILKTFLPELPEPLLTFDLYPHVVGFLNIDESQPVPATLQVLQTLPEENYQVLP
                     FLTAFLVQISAHSDQNKMTNTNLAVVFGPNLLWAKDAAITLKAINPINTFTKFLLDHQ
                     GELFPSPDPSGL"
     misc_feature    753  .785
                     /note="SH3-binding domain; capable of binding the SH3
                     domains of src and p85 alpha; aa residues 251-261"
     misc_feature    849. .926
                     /note="active site box 1; aa residues 283-308"
     misc_feature    999  .1091
                     /note="active site box 2; aa residues 333-363"
     misc_feature    1149. .1277
                     /note="active site box 3; aa residues 383-425"
     polyA_signal    3308. .3313
     polyA_site      3345
                     /note="38 A residues"
BASE COUNT      655 a   1061 c    955 g    774 t
ORIGIN


  Query Match             9.2%;  Score 102.8;  DB 9;  Length 3345;
  Best Local Similarity   54.9%;  Pred. No. 3e-11;
  Matches  203;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy     605 gcagcctcccgactaccactacatcatggagaacctcttcctgtacgtcatcagcagct 664
           | | | |||| | ||||   | |      || ||| |   |||| | |    || |
Db     346 GAATGCCCCCCAGCCACCAGCTCGACCACAAGCAAGCTCCTGGGGTACCTGAAGCACACCC 405

Qy     665 taggagctcctggtggctgaggactacatgatcgtgtacctgaacggtgccacgcccccggc 724
           | || |    ||||    |||||||    |  |||| ||| ||   | |  | || |
Db     406 TGGACCAGTACGTGGAGAAGTGACTACACACTTCTGTATCTGCACCACGGCCTGACCAGCG 465

Qy     725 ggaggatgcctggaatcggctggctgaagaagtgctaccagatgatcgaccggaggttgc 784
             | | |||     || ||||||||     |   ||||| |  | | |||||  | ||
Db     466 ACAACAAGCCCTCCCTCAGCTGGCTCCGTGATGCCTACCGGGAGTTTGACCGCAAGTACA 525

Qy     785 ggaaaaacctgaagtccttgatcatcgtcgatccctctgtggttcattcggactgtgctgg 844
           | ||| ||| | ||| |||||  |||||  | ||  |   ||||||     ||| ||||
Db     526 AGAAGAACATCAAGGCCTTGTACATCGTCCATTCAACCATGTTCATCAAAACTCTGCTCA 585

Qy     845 ccatctctcgccctttcatcagcgtcaaagttcatcaacaagatccagtacgtgcacagct 904
           | |||    ||  |||||||| |||| ||    | ||||||   || ||| |    |
Db     586 TCCTCTTCAAGCCCCTCATCAGCTTCAAGTTCGGGCAGAAGATCTTCTATGTGAATTACC 645

Qy     905 tggaagacctggagcaactcatcctatggaacacgtccagatcccagactgcgtcctgc 964
           ||   || |||    | | | |      ||||  |  |  ||||||   |   || ||
Db     646 TGAGCGAGCTGAGCGAGCACGTGAAGCTCGAGCAGCTGGGGATCCCTCGCCAAGTGCTCA 705

Qy     965 aatacgaaga 974
           |||| || ||
Db     706 AATATGACGA 715
```

```
RESULT   8
AC009556
LOCUS       AC009556   169798 bp    DNA                     HTG       ... NOV-1999
DEFINITION  Homo sapiens chromosome 13 clone RP11-521 M21 map 13, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 13, clone RP11-521M21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Bickerton,S., Bruno,V., Brown,A.,
            Castle,A., Corry,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,D., Dermitzakis,E., Doucet,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,J.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liao,C., Locke,P., Macdonald,P.,
            Marquis,N., McEwan,P., McCurk,A., McKernan,K., McLaughlin,C.,
            Meldrim,J., Molla,M., Morris,W., Murray,L., Mwaluka,F.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,J., Vassiliev,H., Vu,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Yu,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999, Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.genome.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 521M21
            -------------- Summary Statistics
                Sequencing vector: M13, M77815, 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136985 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165100; agarose-tp
                Insert size: 169799, stdev=contigs
                Quality coverage: 4.2 in Q20 bases; agarose-tp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *    1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *    2037     3371: contig of 1335 bp in length
```

```
            *                    gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *                    gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *                    gap of unknown length
            *     5982     9008: contig of 2027 bp in length
            *                    gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *                    gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *                    gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *                    gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *                    gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *                    gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *                    gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *                    gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *                    gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *                    gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *                    gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *                    gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                    gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                    gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                    gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                    gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                    gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                    gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                    gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                    gap of unknown length
            *   153844   169798: contig of 15955 bp in length
FEATURES             Location/Qualifiers
     source          1. .169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a   34914 c   34487 g   49065 t      2 others
ORIGIN


  Query Match             8.3%;  Score 33.6;  DB 44;  Length 169798;
  Best Local Similarity   65.0%;  Pred. No. 1e-10;
  Matches  147;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     199 ...                                                    199
Db  119829 CCTCCTAAATCCTCTGGATCTTAATGACACTCACTCCTGGAGAATCAAGCTCACACGCCCCA 119887

Qy     199 gagatcaacatttctctggatcagagtgaggagtccctgctgtccgatgacttcttggat 258
           |||||  ||||||||||| || |||| |  ||  | || || |||||   ||||||
Db  119888 AATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGATAACTTGGAC 119947

Qy     259 accctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccgactcg 318
           |  || |||||  | || || || ||||||||  | || ||||||||||||  | || ||
Db  119948 AGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCCGATGAAGCAGATTCT 120007
```

```
Qy     319 ctggagttcctggggaatggcaacgaactggagtggaagacgaca 364
           | |||| |  ||  |||| ||  |||  ||| ||| | | |
Db  120008 TTTGAGTACACTGGCCATGGTAAGCCACTAAGTTGGCAAGGCCAAA 120053


RESULT   9
AC014963
LOCUS       AC014963   126121 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, 1 unordered
            pieces.
ACCESSION   AC014963
VERSION     AC014963.1  GI:6436372
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera, Endopterygota, Diptera, Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 126121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211484 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .126121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    36005 a   26506 c   26872 g   36738 t
ORIGIN


  Query Match             8.1%;  Score 89.8;  DB 43;  Length 126121;
  Best Local Similarity   54.4%;  Pred. No. 9.3e-09;
  Matches  203;  Conservative    0;  Mismatches  167;  Indels    3;  Gaps    1;

Qy     644 tcctgtacgtcatcagcagcttagaggctcctggtggctgaggactacatgatcctatacg 703
           ||  | |  ||||||     | || |   | ||||   | ||||||||  | || |||
Db  17312 TCAGGGAAATCATCAAGGAGATCGAACCTTTTGTGGAGAACGACTACATTTIGGTTTACT 17371

Qy     704 tgaacggtgccacgccccggcggaggatgtcctggaatcgactggttgaaggaatgctacc 763
           | |   | |       |  | | ||      |  |||  ||| | |||
Db  17372 TCCATCAAGGCCTCAAGGAGGACAACAAACTATCCCACAATTTTTTTTTTTTTTTTTCC 17431

Qy     764 agatgatcgacccggaggttgcggaaaaacctgaaagtccttgatcatcgtccgccccctcct 823
           ||  | |||| || |  || || || || |||||| | |||   | || || ||  |||
Db  17432 AGGAGCTCGATCGCAACTTCCGCAAGAATCTGAAGACGTTGTATGTGGTTCATCCGACGT 17491

Qy     824 ggttcattcggactgtgctggcgatatctgcgcctttcatcaggtgaagttcatcaa 883
           ||||||| |||    |   | ||| | |||| ||||| |||| ||||||  ||| .
Db  17492 GGTTCATACGGGTCATTTGGAACTTCTTTAGCCCGTTCATTAGCGACAAGTTCCGCAAGA 17551

Qy     884 agatccagtacgtgcacagcttggagagcctggagcaactcatcctctgtgtc atc c 943
           || |    ||| |  |||||||| || ||  ||     ||    || |  | |
Db  17552 AGCTAGTCTACATCTCCAGCTTGGACGAGTTGCGCCAGCGCTCCGCCTGAACAAGCTCA 17611

Qy     944 agatcccagactgcgtcctgcaat.tgtttttgaagctgatgagttgtgtttttttt a 1003
           || | || ||| | ||      ||  || || | ||||||  |||| ||||
Db  17612 AGTTCCCGGACAACATC---TGCAGCTTCGACGACAAACTGAACCCCAGCCCCAACCCAT 17668

Qy    1004 cgaggccccagcc 1016
           |||  || | |||
Db  17669 CGACACCGCCGCC 17681


RESULT  10
AC009556/c
```

```
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone PP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone PP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,P., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depaire,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,M., Morrow,J., Mychalecky,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,P., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1016. contig of 1016 bp in length
            *                   gap of unknown length
            *    1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *    2037     3371. contig of 1335 bp in length
            *                   gap of unknown length
            *    3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *    4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *    5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *    8009     9286. contig of 1278 bp in length
            *                   gap of unknown length
            *    9287    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *   11244    14312. contig of 3069 bp in length
            *                   gap of unknown length
            *   14313    18423. contig of 4111 bp in length
            *                   gap of unknown length
            *   18424    23065. contig of 4642 bp in length
            *                   gap of unknown length
            *   23066    30558. contig of 7493 bp in length
            *                   gap of unknown length
            *   30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *   37538    44831. contig of 7294 bp in length
            *                   gap of unknown length
            *   44832    50944. contig of 6113 bp in length
            *                   gap of unknown length
            *   50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *   58264    65216. contig of 6953 bp in length
            *                   gap of unknown length
            *   65217    72310. contig of 7094 bp in length
            *                   gap of unknown length
            *   72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *   78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *   87608    98337. contig of 10730 bp in length
            *                   gap of unknown length
            *   98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *  112199   126330. contig of 14132 bp in length
            *                   gap of unknown length
            *  126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *  140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *  153844   169798: contig of 15955 bp in length
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="PP11-57C21"
                     /clone_lib="PPCI-11 Human Male BAC"
BASE COUNT    51330 a   34914 c   34457 g   49097 t       2 others
ORIGIN
```

```
  Query Match             7.5%;  Score 88.6;  DB 44;  Length 169798;
  Best Local Similarity   75.4%;  Pred. No. 1.6e-07;
  Matches  104;  Conservative    0;  Mismatches   34;  Indels     1;  Gaps     0;

Qy     644 tcctgtacgttcatcagcagcttaaagatcttttgtgctgagga ta atgatttttaa 703
           ||  || || || || | ||                                       |
Db   13166 TCAGATATGTAATAAGTACTTTAGAGTTGATGGTAGCTGAAGATTATATGAATTGTGTACT 13107

Qy     704 tgaacggtgccacgccccgggcgaggatccctgaatggctcctcaagaaatactgcc 760
           |||| ||||| ||| |  |                                      |
Db   13106 TGAATGGTGCAACCCCAAGAAGGAGGATGCCAGGGTAGGTGAAAAAAGTATAT 13047

Qy     764 agatgatcgaccggaggt 781
           |||||| || |  |  ||
Db   13046 AGATGATTGACAGACGGT 13029

RESULT  11
```

```
AC015396/c
LOCUS       AC015396    78170 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015396
VERSION     AC015396.1  GI:6435939
KEYWORDS    HTG, HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78170)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209602 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/qualifiers
     source          1   78170
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21762 a  17118 c  17312 g  21978 t
ORIGIN


  Query Match             7.2%;  Score 80.4;  DB 43;  Length 78170,
  Best Local Similarity   59.0%;  Pred  No. 7.2e-07;
  Matches  157;  Conservative    0;  Mismatches  106;  Indels    3;  Gaps    1;

Qy     496 atagacctgcacatgatccggcctacatgaaagtggtcacccacggaggga---ctac 552
           || ||| |||   | ||   ||| |||| |   ||| |  ||||||| || ||   | |
Db    8691 ATCGACATGCGGGTCATAGAGCCCTACAAGGTGTGCTGTCCCACGGCGGCTATCTCAAA 8632

Qy     553 ggcgaaggcctcaacgccatcatcgtcttcgcagcctgcttccttccagacagcagcctc 612
             |  || |  ||||||||  || ||||   |||||   ||| || || ||
Db    8631 ICTGGCGGICAGAACGCCATTGTCATCTTTIGCGCCTGTCACCTGCCGGATAGGTCGAGG 8572

Qy     613 ccagactaccactacatcatggagaacctcttcctgtacgtcatcagcagcttagagctc 672
           ||||||||  |||| | ||||| ||  | | |  ||||||||  |||  | | | || |
Db    8571 GCCGACIACAGCIACGIGATGGACAATTIGTIICTGTACGTIGTCAAGACCCTGGAACAG 8512

Qy     673 ctggtggctgaggactacatgatcgtgtacctgaacggtgccacgccccggcggaggatg 732
           ||||  |||| |||||  || |  | || ||' | || | | |     |||| |   ||
Db    8511 TTGGTCACTGATGACTATGTGCTAATGTATTTCATGGGGGCTCCAATAGGCGTAATGTG 8452

Qy     733 cctggaatcggctggctgaagaagtg 758
           ||     |  ||||| ||   |||
Db    8451 CCACCTTTTGCCTGGCTTAAACGGTG 8426


RESULT  12
AC011488
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *                     gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *                     gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *                     gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *                     gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *                     gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *                     gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *                     gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *                     gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *                     gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *                     gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *                     gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *                     gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *                     gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *                     gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *                     gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *                     gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *                     gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *                     gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *                     gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *                     gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *                     gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *                     gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *                     gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *                     gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *                     gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *                     gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *                     gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *                     gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *                     gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *                     gap of unknown length
```

```
            *    24908    25839: contig of 932 bp in length
            *                   gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *                   gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *                   gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *                   gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *                   gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *                   gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *                   gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *                   gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *                   gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *                   gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *                   gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *                   gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *                   gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *                   gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *                   gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *                   gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *                   gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *                   gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *                   gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *                   gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *                   gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *                   gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *                   gap of unknown length
            *   112601   179301: contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42599 t     85 others
ORIGIN


  Query Match             6.9%;  Score 76.4;  DB 41;  Length 179301;
  Best Local Similarity   98.7%;  Pred. No. 4.3e-06;
  Matches   77;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     999 agaagcgcagaggcccagcccggagtttgtgtgtgcccaggtctgaagagaagccagaggtgg 1057
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   83709 ACAGCCCGAGGCCCCAGCCCGGAGTTTGTGCTGCCCAGGTCTGAAGAGAAGCCAGAGGTGG 83768

Qy    1058 caccagtggaaaacaggt 1075
           ||||||||||||||||||
Db   83769 CACCAGTGGAAAACAGGT 83786


RESULT  13
HSU62794
LOCUS       HSU62794     1378 bp    mRNA            PRI       13-AUG-1997
DEFINITION  Human CDC42 GAP-related protein mRNA, partial cds.
ACCESSION   U62794
VERSION     U62794.1  GI:2326170
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 269 (35), 26650-26653 (1994)
  MEDLINE   94375275
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (1), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   3  (bases 1 to 1378)
  AUTHORS   Barfod,E.T. and Ashkenazi,A.
  TITLE     An effector function for Cdc42GAP that modulates the mitotic
            spindle
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Chan,M., Kuang,W.J. and Ashkenazi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1996) Molecular Oncology, Genentech, 460 Point
            San Bruno Blvd., S. San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1378
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CMK"
                     /cell_type="megakaryocyte"
     CDS             <1..672
                     /note="CDC42GAP; similar to CDC42 GTPase activating
                     protein encoded by GenBank Accession Number D28570;
                     insertion in the rhoGAP active site renders active site
                     boxes 2 and 3 out of frame and untranslated"
                     /codon_start=1
                     /product="CDC42 GAP-related protein"
                     /protein_id="AAB66594.1"
                     /db_xref="GI:2326171"
                     /translation="SSNLLRSHSSLRLGLSLSKKKSMHWIRSVAPVPPRLLLSCPFCL
                     CPYFKNIKALYVHRTMFDKTLTLFKPLISPKPDQLKIPVVNVLSPLSEHVKLEQLGI
                     PRIVLPYDFLPSIPSPRTAPMFRPPLPNLLPCVSLLKLSRNFPLEPIPIVLR
                     ETVAYLQAHALTTEGIFRRSANTQVVREVLQKYNMLGCPSGCARLLSPCPHPRAACG
                     FRPVQ"
     misc_feature    370..402
                     /note="partial 5' [illegible] 31 g [illegible] capable of binding src
                     and p85"
     misc_feature    465..543
                     /note="active site box 1"
     misc_feature    585..640
                     /note="55 nucleotide insertion into sequence identical
                     with CDC42GAP; insertion between active site boxes 1 and 2
                     shifts translational frame such that a stop codon occurs
                     at 670, and active site boxes 2 and 3 are not made"
     misc_feature    671..763
                     /note="out-of-frame Anti-rho site box 2"
     misc_feature    821..949
                     /note="Anti-rho site box 3"
     polyA_site      1378
                     /note="21 A nucleotides"
BASE COUNT      288 a    468 c    328 g    294 t
ORIGIN


  Query Match             6.1%;  Score 68.2;  DB 1;  Length 1378;
```

```
  Best Local Similarity   58.7%;  Pred. No. 0.00026;
  Matches  118;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy    774 cggaggttgcggaaaaacctgaagtccttgatcatcgtccacccctcgtggttcattcg 833
          | | ||||    ||| ||| | ||| |||||  |||||| || ||  |   ||||||
Db    132 CTGCAGGTACAAGAAGAACATCAAGGCCTTGTACATCGTGCATCCAACCATGTTCATCAA 191

Qy    834 gactgtgtgtggcgatctatgcgccttttcatcaacgtcaagttcatcaacaagatccagta 893
          ||| |||  | |||    ||  |||||||| ||||||||    | ||||||   ||
Db    192 AACTCTGTTCATCCTCTTCAAGCCCCTCATCACGTTCAAGTTCGGGCAGAAGATCTTCTA 251

Qy    894 cgtgcacagctttggaagacctggagcaactcatccctatggaacacgtccagatcccaga 953
          ||| | |  | ||   || |||   | | | |    |||| || |   ||||||
Db    252 TCTCAATTACCTGAGCGACTGAGTGACTACTCGAAGTTGGAGCAGCTGGGGATCCTTCG 311

Qy    954 ctgctctgcaatacgaaga 974
          |   || ||  |||| || ||
Db    312 CCAAGTCGTTCAAATATGACGA 332


RESULT  14
AB027763/c
LOCUS       AB027763     6573 bp    DNA             BCT       01-JUN-1999
DEFINITION  Streptomyces fradiae gene, partial cds.
ACCESSION   AB027763
VERSION     AB027763.1  GI:4958945
KEYWORDS    .
SOURCE      Streptomyces fradiae (strain:Y59) DNA.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales, Streptomycineae, Streptomycetaceae; Streptomyces
REFERENCE   1  (bases 1 to 6573)
  AUTHORS   Ogawara,H.
  TITLE     Streptomyces
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 6573)
  AUTHORS   Ogawara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1999) to the DDBJ/EMBL/GenBank databases. Hiroshi
            Ogawara, Meiji Pharmaceutical University, Department of
            Biochemistry; Noshio,2-522-1, Kiyose, Tokyo 204-8588, Japan
            (E-mail:hogawara@my-pharm.ac.jp, Tel:81-424-95-8474,
            Fax:81-424-95-8474)
FEATURES             Location/Qualifiers
     source          1. .6573
                     /organism="Streptomyces fradiae"
                     /strain="Y59"
                     /db_xref="taxon:1806"
     CDS             complement(<1. .1310)
                     /note="product is unkown; the 600th nucleotide of this
                     sequence is corresponding to the 1st nucleotide of the
                     sequence of acc#AB024056"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA78104.1"
                     /db_xref="GI:4958946"
                     /translation="MAPPAPDSAVGVTGEPFAGTDAQGPVPGFVDAHNHIMSNEAFGG
                     RIIQGKPFSERGVADAIDCPHHYPDGSLALFDFITPGGDGRHDPDGWPTFADWPAHD
                     SITHQQNYYAWIERALARGQRVLVNDLVTNGVICSVYFFKDRGCDEMESIPLQARPTY
                     EMQEYVDRMYGGPGKGWPFINTDSAQAREVVPQGPLAVVLGVETSEPFGCKQVLDVAR
                     CDREDIDRGLDELHRLGVRSMFLCHKFDNALCGVPFDSGALGTAVNVGQFLSTGTFWP
                     IEPCTGPQHDNPIGGASAPGAEAKLPPGVTVPAYAEGARCNTRGLIDLGAYAVPGMMP
                     RGMMLFIDHMSVKAAGRAIDIIESEAYPGVISSHSWMDAEWTEPVYRLGGFIAQYMHG
                     SAQFTAEARPIDALPAPYGVGYGYGTDMNGVGGWPAPPGADAPDP"
BASE COUNT      770 a   2395 c   2685 g    722 t
ORIGIN


  Query Match             4.9%;  Score 54.4;  DB 1;  Length 6573;
  Best Local Similarity   45.8%;  Pred. No. 0.13;
  Matches  182;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy    290 acatggaggaccccgatgagaccgactactgttggagttcctggggaatggcaacgaactgg 349
          |||||  | ||  ||| | |  || |  ||    | ||   |||||   | |
Db   1207 ACATCATGTCCAACGAGGCGTTCGGCGGCGCCTGATCTGCGGGAAGCCCTTCTCCGAAC 1148

Qy    350 agtgggaagacgacacccccgtggccacccgccaagaacatgcccggggacagcggcggatc 409
          | | | |||| ||| || || |    || || ||   ||||  | |  ' ||  :
Db   1147 GCGGCGTCGCCGACGCCCTCGAGGACTGTCCCGAGCACTACCCCGACGGCTCCCTGGCGT 1088

Qy    410 tatttggggacggcacgacggaggacggcagcggcggccaacggcgccctgtcccgcgcgcag 469
          | || |   | ||| | || | || | || | ||  | |  |  |  ||||  ||
Db   1087 TCTTCGACTTCATCACCAGGGGCGGCCACGGCAGGCACGACCCCGACGGCTGGCCCACCT 1028

Qy    470 tgatcatcggggagcaagagcacggtatagacctgcacatgatccggccttacatgaaag 529
          |   |  | ||     |  ||    |  ||  || ||  |  |  |  |  |  |  |
Db   1027 TCGCCGACTGGCCGCCCACGACATCCCTCAGCTACCAGCAGAAGTACTACCCCTCCATCG 968

Qy    530 tggtcacccacggaggggtacta                                    589
          |  | |||  |  |    |  |||    | |||||| | || ||  |  |  | |  |
Db    967 AGCGCGCCCTGGCGCGCGGGCAGCGCGTCCTCGTCAACGACGCTCGTCAGCAACGGCGTA 906

Qy    590 gcttccttccagacagcagcctccccgactaccactacatcatggaagaacctcttcctat 649
          || |       ||  |  |      |  |  |  | ||   ||||||  | ||  |||
Db    907 TCTGCTCGGTCTACTTCTTCAACGACGCTCGTCGTCAACGACGCTCGTCAGCAAGACGTA 848

Qy    650 acgtcatcagcagcttagagctcctggtggctgaggaactacatgatcg 697
          | | |  ||| |  |  ||| | | || |   | ||||  ||||  ||
Db    847 AGGCCCGCAGGACGTACGAGATGCAGGAGTACGTGGACCGCATGTACG 800


RESULT  15
AR029509
LOCUS       AR029509     3624 bp    DNA             PAT       24-SEP-1999
DEFINITION  Sequence 6 from patent US 5859336.
ACCESSION   AR029509
VERSION     AR029509.1  GI:5941492
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3624)
  AUTHORS   Koziel,M.G., Desai,N.M., Lewis,K.S., Warren,G.W., Evola,S.V.,
            Crossland,L.D., Wright,M.S., Merlin,E.J., Lennis,K.L., Bowman,C.G.,
            Dawson,J.L., Dunder,E.M., Pace,G.W. and Suttie,J.L.
  TITLE     Synthetic DNA sequence having enhanced activity in maize
  JOURNAL   Patent: US 5859336-A 6 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .3624
                     /organism="unknown"
BASE COUNT      764 a   1347 c   1061 g    452 t
ORIGIN


  Query Match             4.7%;  Score 52;  DB 5;  Length 3624;
  Best Local Similarity   45.7%;  Pred. No. 0.41;
  Matches  181;  Conservative    0;  Mismatches  215;  Indels    0;  Gaps    0;

Qy    268 gacctggatattaacgtggatgacatcgaggacccccgatgaggacggactctctagggttc 327
          | |||| |  | || |    || | ||||   |||  ||||||  |  ||
Db   1249 GGCCTGCAGCTGAAGGACAGCGAGACCCAGCTGCCCCCCGAGACCACCGAGGCCCCCAAT 1308

Qy    328 ctggggaatggcaacgaactgggg                                   387
          | ||    || | | |    |    |  || || | | ||  : :    ||
Db   1309 TACGAGAGCTACAGCCACCGCCTGACCCACATCGGCATCATCTTCTCCAGAGCCCGCTGAAC 1368

Qy    388 atgcccgggacagcgcgggatc                                     447
          ||||||  |||||   |    |  | || |  |  |  |  |  |  |  |  |
Db   1369 GTGCCCGTGTACAGCTGGACCCACCGCAGCGCCGACCGCACCAACACCATCTGCCCCAAC 1428

Qy    448 aacgggcgcctgtggcggacagtgatcatcgggtgagcaagagcaccgtatagaccttcgg 507
          |  || |  | | |  ||||    | | ||||    ||  | |    |  ||
Db   1429 CGCATCACCCAGATCCCCATGGTGAAGGCCAGCGACCTGCCCCAGGGCACCACCGTGGTG 1488

Qy    508 atgatccggccttacatgaaagtggtcacccacggagggtactaccggcgaaagcctcaac 567
            ||    | ||    |  | || ||   |  |   | ||  ||  ||| ||  |
```

```
Db   1489 CGCGGCCCCGGCTTCACCGGCGGCGACATCC:GCGCCGCACCAACACCGGCGGCTTCGGC 1548

Qy    569 gccatcatcgtcttcgcagcctgcttccttccagacagcagcctcccgactaccactac 627
          ||||| ||| || | || ||| | || || | || || || ||||
Db   1549 CCCATCCGCGTGACCGTGAACGGCCCCCTGACCCAGCGCTACCGCATCGGCTTCCGCTAC 1608

Qy    628 atcatggagaacctcttcctgtacgtcatcagcagc 663
            ||          |||| | | || | | || ||
Db   1609 GCCAGCACCGTGGACTTCGACTTCTTCGTGAGCCGC 1644


Search completed: February 9, 2000, 07:12:44
Job time: 10078 sec
```

```
M of: US-09-174-937-2 to: GenEmbl:*  out_format : pfs

Date: Feb 9, 2000  4:33 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c, 1993 1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV-xlp
 -Q=/cgn2_1/"SPTO_spool/US09174937/runat_03022000_095149_13857/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCT=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -SIAKT=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block.
Query: US-09-174-937-2
Query length: 371
Database: GenEmbl:*
Database sequences: 821193
Database length: 1518192014
Search time (sec): 1186.310000

WARN: XGAPOP and YGAPOP must be equal  Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


Score_list.
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_htg3:AC011488    -    73.00 1288.79  9.2e-64   179301  ! AC011488 Homo sapiens chrom
gb_htg3:AC011488    -    25.00  402.41  2.2e-14   179301  ! AC011488 Homo sapiens chromo
gb_pat:AP028453     -    23.00  396.09  4.9e-14     1386  ' AP028453 Sequence 22 from pate
gb_pr3:HS115179     -    23.00  392.69  7.7e-14     2382  ' Z115179 Homo sapiens BCL2/XL
gb_ro:AF035207      -    23.00  391.14  9.1e-14     2213  ! AF035207 Mus musculus Nip1 mR
gb_htg6:AC009556    -    21.00  329.89  2.7e-10   169798  ! AC009556 Homo sapiens chrom
gb_pr1:AB002365     -    17.00  276.44  1.2e-07     5654  ! AB002365 Human mRNA for KIAA0
gb_pl1:SCN200XG     -    11.00  162.80  0.4799      8987  ! X77114 S.cerevisiae N2000, N20
gb_pl1:SCYNL005C    -    11.00  171.16  0.1642      2353  ' Z71281 S.cerevisiae chromosome
gb_pl1:YSCMRP7      -    11.00  172.14  0.1449      2016  ! M22110 Yeast (S.cerevisiae) mi
gb_vi:AF123485      -     9.00  130.10  22.67       4513  ! AF123485 San Angelo virus pol
gb_htg3:AC009344    -     9.00  109.82  428.65    113791  ! AC009344 Drosophila melanoga
gb_htg3:AC009348    -     9.00  109.55  501.67    139296  ! AC009348 Drosophila melanoga
gb_htg3:AC010754    -     9.00  106.69  640.37    187096  ! AC010754 Homo sapiens chromo
gb_ba2:AP005404     -     8.00   88.41  6.7e-03   181654  ! AP005404 Bacillus anthracis
gb_ba2:BFU63096     -     8.00  108.23  525.34      7786  ! U63096 Bacteroides fragilis (
gb_ba2:BSU00819     -     8.00  104.27  673.42     14616  ! U00819 Bacillus subtilis spa
gb_ba2:REPCOBLMK    -     8.00  110.14  421.49      5753  ! L21196 Rhodococcus erythropoli
gb_ov:GGU67088      -     8.00  119.91  117.52      1218  ! U67088 Gallus gallus T-box con
gb_ov:MIGCYP1G      -     8.00  109.05  473.01      6837  ! L41917 Microgadus tomcod cytoc
gb_pl1:AB011482     -     8.00   93.30  3.6e-03    83478  ! AB011482 Arabidopsis thaliana
gb_pl1:PR5SRRN2     -     8.00  122.17   87.99       851  ! Z33604 P.radiata (Spr10) 5S rR
gb_pl1:SPBC17D1     -     8.00  102.30  1.1e-02    20000  ' AL021322 S.pombe chromosome I
gb_pl1:SPCC191      -     8.00  100.89  1.3e-02    25000  ' AL049644 S.pombe chromosome I
gb_pl2:AC000103     -     8.00   92.21  4.1e-03    99383  ' AC000103 Genomic sequence for
gb_pl2:AC000294     -     8.00   91.96  4.5e-03   102339  ' AC000294 Arabidopsis thalian
gb_pl2:AC004255     -     8.00   92.31  4.1e-03    97799  ! AC004255 Arabidopsis thaliana
gb_pl2:AC009243     -     8.00   91.59  4.5e-03   110565  ! AC009243 Genomic sequence fo
gb_pl2:AMA250397    -     8.00  120.33  111.40      1140  ! AJ250397 Antirrhinum majus par
gb_pl2:ATAC004667   -     8.00   92.70  3.9e-03    91863  ! AC004667 Arabidopsis thaliana
gb_pl2:ATAC006264   -     8.00   92.61  4.0e-03    91222  ! AC006264 Arabidopsis thaliana
gb_pl2:T4B21        -     8.00   92.79  3.8e-03    91477  ! AF118223 Arabidopsis thaliana
gb_pr1:HS164D6F     -     8.00  132.42  23.63        167  ! Z59580 H.sapiens CpG island DNA
gb_pr2:AB033059     -     8.00  110.99  371.70      5072  ! AB033059 Homo sapiens mRNA for
gb_pr2:AP000131     -     8.00   92.17  4.1e-03   100000  ! AP000131 Homo sapiens genomi
gb_pr2:AP000209     -     8.00   92.17  4.1e-03   100000  ! AP000209 Homo sapiens genomi
gb_pr2:AP000247     -     8.00   91.29  4.7e-03   115119  ! AP000247 Homo sapiens genomi
gb_pr2:CNS01BOV     -     8.00   87.49  7.5e-03   210134  ! AL109758 Human chromosome 14
gb_pr2:HS134P22     -     8.00   91.07  4.7e-03   118951  ! AL035403 Human DNA sequence
gb_pr2:HS200328     -     8.00   [illegible]  5.0e-3   1.7[illegible]  ! AL 2[illegible] Human DNA seque
gb_pr2:HS207P19     -     8.00   91.96  4.5e-03   [illegible]  ! [illegible] Human DNA sequenc
gb_pr2:HS213C16     -     8.00   [illegible]  5.0e-3   [illegible]  ! AL [illegible] Human DNA seque
gb_pr2:HSY12298     -     8.00   [illegible]  [illegible]  [illegible]  ! [illegible] Homo sapiens GDP di
gb_pr3:AC004014     -     8.00   88.07  6.4e-3   172441  ! AF [illegible] Human BAC clone
gb_htg7:AC010606    -     8.00   [illegible]  5.2e-3   [illegible]  ! [illegible] Homo sapiens cl

seq_name: gb_htg3:AC011488

seq_documentation_block:
LOCUS       AC011488   179301 bp    DNA                 HTG  [illegible]-FEB-1999
DEFINITION  Homo sapiens chromosome [illegible]  [illegible]  *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015284
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a "working draft" sequence.  It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary.  Gaps between the contigs are represented as
            * runs of N, but the length of the gaps is only
            * approximate.  As soon as it is possible to determine the
            * true order of the contigs, they will be submitted again
            * and the new sequence record will supersede this one.
            * The working draft will be updated until the finished
            * sequence is released.  Please check back frequently
            * for improvements to the sequence.  Keywords, references,
            * and the table of features will be added when the finished
            * sequence is released.  Until then, please cite the
            * publication listed in the publications section of the
            * HTGS publication list.
            * The sequence in this record is derived from clones
            * RP11-[illegible], RP11-[illegible], RP11-[illegible], RP11-[illegible], RP11-[illegible],
```

```
*                 gap of unknown length
*   13948   15016: contig of 1069 bp in length
*                 gap of unknown length
*   15017   15614: contig of 598 bp in length
*                 gap of unknown length
*   15615   16687: contig of 1073 bp in length
*                 gap of unknown length
*   16688   18152: contig of 1465 bp in length
*                 gap of unknown length
*   18153   19221: contig of 1069 bp in length
*                 gap of unknown length
*   19222   20477: contig of 1256 bp in length
*                 gap of unknown length
*   20478   21388: contig of 911 bp in length
*                 gap of unknown length
*   21389   22058: contig of 670 bp in length
*                 gap of unknown length
*   22059   23154: contig of 1096 bp in length
*                 gap of unknown length
*   23155   23925: contig of 771 bp in length
*                 gap of unknown length
*   23926   24290: contig of 365 bp in length
*                 gap of unknown length
*   24291   24907: contig of 617 bp in length
*                 gap of unknown length
*   24908   25839: contig of 932 bp in length
*                 gap of unknown length
*   25840   27116: contig of 1277 bp in length
*                 gap of unknown length
*   27117   27904: contig of 788 bp in length
*                 gap of unknown length
*   27905   28519: contig of 615 bp in length
*                 gap of unknown length
*   28520   28930: contig of 411 bp in length
*                 gap of unknown length
*   28931   30384: contig of 1454 bp in length
*                 gap of unknown length
*   30385   31245: contig of 861 bp in length
*                 gap of unknown length
*   31246   32148: contig of 903 bp in length
*                 gap of unknown length
*   32149   33058: contig of 910 bp in length
*                 gap of unknown length
*   33059   34020: contig of 962 bp in length
*                 gap of unknown length
*   34021   34782: contig of 762 bp in length
*                 gap of unknown length
*   34783   36041: contig of 1259 bp in length
*                 gap of unknown length
*   36042   37033: contig of 992 bp in length
*                 gap of unknown length
*   37034   37565: contig of 532 bp in length
*                 gap of unknown length
*   37566   38344: contig of 779 bp in length
*                 gap of unknown length
*   38345   39359: contig of 1015 bp in length
*                 gap of unknown length
*   39360   41242: contig of 1883 bp in length
*                 gap of unknown length
*   41243   41839: contig of 597 bp in length
*                 gap of unknown length
*   41840   42836: contig of 997 bp in length
*                 gap of unknown length
*   42837   43913: contig of 1077 bp in length
*                 gap of unknown length
*   43914   55087: contig of 11174 bp in length
*                 gap of unknown length
*   55088   81867: contig of 26780 bp in length
*                 gap of unknown length
*   81868  112600: contig of 30733 bp in length
*                 gap of unknown length
*  112601  179301: contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  4747[illegible] g  42598 t     85 others
ORIGIN

alignment_scores:
            Quality:   73.00           Length:     73
              Ratio:   1.000             Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC011489/rev   ..

 Align seg 1/1  to reverse of: AC011488  from: 1  to: 179301

      47 ProProAsnThrLeuAsnPheAs[illegible]LysThrLe 63
         |||||||||||||||||||||||||||||||||||||||||||||||||
  124301 CCTCCCAACACGCTAAATTTCAA[illegible]AAGACGCT 124252

      63 uValAlaProGluIleAsnIleS[illegible] 80
         |||||||||||||||||||||||||||||||||||||||||||||||||
  124251 GGTGGCCCCAGAGATCAACATTT[illegible] 124202

      80 euSerAspAspPheLeuAspThr[illegible]Asp 96
         |||||||||||||||||||||||||||||||||||||||||||||||||
  124201 TGTCCGATGACTTCTTGGATAC[illegible]GGATATTAA[illegible] 124152

      97 AspIleGluThrProAspGluThrAsp[illegible] 11[illegible]
         |||||||||||||||||||||||||||||||||||||||||||||||||
  124151 GACATCGAGACCCCCGATGAGAC[illegible] 124102

     113 yAsnGluLeuGluTrpGlu 119
         |||||||||||||||||||
  124101 CAACGAACTGGAGTGGGAA 124083

seq_name: gb_htg3:AC011488

seq_documentation_block:
LOCUS       AC011488   179301 bp    DNA             HTG       [illegible]-1999
DEFINITION  Homo sapiens chromosome [illegible] *** [illegible]ENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            www.jgi.doe.gov.
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *                  gap of unknown length
            *     1008     1936: contig of 929 bp in length
```

```
*                    gap of unknown length
*       1937      2810: contig of 874 bp in length
*                    gap of unknown length
*       2811      3791: contig of 981 bp in length
*                    gap of unknown length
*       3792      4639: contig of 848 bp in length
*                    gap of unknown length
*       4640      5254: contig of 615 bp in length
*                    gap of unknown length
*       5255      5713: contig of 459 bp in length
*                    gap of unknown length
*       5714      6345: contig of 632 bp in length
*                    gap of unknown length
*       6346      6971: contig of 626 bp in length
*                    gap of unknown length
*       6972      7837: contig of 866 bp in length
*                    gap of unknown length
*       7838      8339: contig of 502 bp in length
*                    gap of unknown length
*       8340      8880: contig of 541 bp in length
*                    gap of unknown length
*       8881      9498: contig of 618 bp in length
*                    gap of unknown length
*       9499     10379: contig of 881 bp in length
*                    gap of unknown length
*      10380     10992: contig of 613 bp in length
*                    gap of unknown length
*      10993     11901: contig of 909 bp in length
*                    gap of unknown length
*      11902     13352: contig of 1451 bp in length
*                    gap of unknown length
*      13353     13947: contig of 595 bp in length
*                    gap of unknown length
*      13948     15016: contig of 1069 bp in length
*                    gap of unknown length
*      15017     15614: contig of 598 bp in length
*                    gap of unknown length
*      15615     16687: contig of 1073 bp in length
*                    gap of unknown length
*      16688     18152: contig of 1465 bp in length
*                    gap of unknown length
*      18153     19221: contig of 1069 bp in length
*                    gap of unknown length
*      19222     20477: contig of 1256 bp in length
*                    gap of unknown length
*      20478     21388: contig of 911 bp in length
*                    gap of unknown length
*      21389     22058: contig of 670 bp in length
*                    gap of unknown length
*      22059     23154: contig of 1096 bp in length
*                    gap of unknown length
*      23155     23925: contig of 771 bp in length
*                    gap of unknown length
*      23926     24290: contig of 365 bp in length
*                    gap of unknown length
*      24291     24907: contig of 617 bp in length
*                    gap of unknown length
*      24908     25839: contig of 932 bp in length
*                    gap of unknown length
*      25840     27116: contig of 1277 bp in length
*                    gap of unknown length
*      27117     27904: contig of 788 bp in length
*                    gap of unknown length
*      27905     28519: contig of 615 bp in length
*                    gap of unknown length
*      28520     28930: contig of 411 bp in length
*                    gap of unknown length
*      28931     30384: contig of 1454 bp in length
*                    gap of unknown length
*      30385     31245: contig of 861 bp in length
*                    gap of unknown length
*      31246     32148: contig of 903 bp in length
*                    gap of unknown length
*      32149     33058: contig of 910 bp in length
*                    gap of unknown length
*      33059     34020: contig of 962 bp in length
*                    gap of unknown length
*      34021     34782: contig of 762 bp in length
*                    gap of unknown length
*      34783     36041: contig of 1259 bp in length
*                    gap of unknown length
*      36042     37033: contig of 992 bp in length
*                    gap of unknown length
*      37034     37565: contig of 532 bp in length
*                    gap of unknown length
*      37566     38344: contig of 779 bp in length
*                    gap of unknown length
*      38345     39359: contig of 1015 bp in length
*                    gap of unknown length
*      39360     41242: contig of 1883 bp in length
*                    gap of unknown length
*      41243     41839: contig of 597 bp in length
*                    gap of unknown length
*      41840     42836: contig of 997 bp in length
*                    gap of unknown length
*      42837     43913: contig of 1077 bp in length
*                    gap of unknown length
*      43914     55087: contig of 11174 bp in length
*                    gap of unknown length
*      55088     81867: contig of 26780 bp in length
*                    gap of unknown length
*      81868    112600: contig of 30733 bp in length
*                    gap of unknown length
*     112601    179301: contig of 66701 bp in length
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="CIT978KB_171N13"
BASE COUNT    42023 a  47125 c  47471 g  42682 t      0 others
ORIGIN


alignment_scores:
            Quality:   25.00              length:     37
              Ratio:  1.000                bars:
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC011488    ..

 Align seq 1/1  to. AC011488  from: 1  to: 179301

     334 SerAlaArgProGlnProGluProValGluProArgSerProProPro 370
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   83711 AGCGCGAGGCCCCAGCCGGAGTTTGTCGAGCCGCGAAAACCCCCCCCC 83674

     350 GluValAlaProValGluAsnArg 376
         |||||||||||||||||||||||||
   83761 AGAGGTGGCACCAGTGGAAAACAG 83738

seq_name: gb_pat:AR028453

seq_documentation_block:
LOCUS       AR028453     1386 bp    DNA             HAT       ,--981-1499
DEFINITION  Sequence 22 from patent US 5874467.
ACCESSION   AR028453
VERSION     AR028453.1  GI:5940416
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1386)
  AUTHORS   Chinnadurai,G.
  TITLE     Apoptosis-regulating p. 7kDa
```

```
  JOURNAL   Patent: US 5858678-A 22 12 JAN 1999,
FEATURES             Location/Qualifiers
     source          1. .1386
                     /organism="unknown"
BASE COUNT      428 a    249 c    336 g    373 t
ORIGIN


alignment_scores:
            Quality:   23.00             Length:     23
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AR028453   ..

 Align seq 1/1  to: AR028453  from: 1  to: 1386

     257 IleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHisPr 273
         ||||||||||||||||||||||||| || |||||||||||||||||||||
     910 ATTGATAGAAGGTTACGGAAAAATCTAAAATCCCTAATCATTGTACATCC 959

     273 oSerTrpPheIleArgThr 279
         |||||||||||||||||||
     960 TTCTTGGTTTATCAGAACA 978

seq_name: gb_pr3:HSU15173

seq_documentation_block:
LOCUS       HSU15173     2382 bp    mRNA            PRI       02-FEB-1999
DEFINITION  Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2)
            mRNA, complete cds.
ACCESSION   U15173
VERSION     U15173.1  GI:558843
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 212 to 1156)
  AUTHORS   Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K.,
            Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G
  TITLE     Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
            of cellular proteins
  JOURNAL   Cell 79, 341-351 (1994)
  MEDLINE   95042730
REFERENCE   2  (bases 1 to 2382)
  AUTHORS   Boyd,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1994) Janice M. Boyd, Institute for Molecular
            Virology, St. Louis University Health Sciences Center, 3681 Park
            Avenue, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B-cell, S Elledge, B-cell, J.Ambrus, fetal
                     brain, M.Green"
                     /cell_type="B cells; brain"
                     /dev_stage="fetus (brain)"
     gene            1. .2382
                     /gene="BNIP2"
     CDS             212. .1156
                     /gene="BNIP2"
                     /codon_start=1
                     /product="BCL2/adenovirus E1B 19kD-interacting protein 2"
                     /protein_id="AAC00002.1"
                     /db_xref="GI:558844"
                     /translation="MEGVELKREWQDEDFFIPLPEDDSIEADILAIIGPEDQPGSLEV
                     NGNKVRKKLMAPDISLILDPSDGSVLSDDLDESGEIDLDGLDTPSENSNEFEWEDDLP
                     KPKTTEVIPKGSITFYTAAPPKEDGPPWPMPPIGPQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQFNYRYLMDNLFKYVIGTLELLVAENYMIVYLNGAITRR
                     KMPSLGWLPKCYQQIDPRIPKNLKSLIIVHPSWFIPTLLAVTPPFISSKFSQKIPYVF
                     NLAELAELVPMEYVQIPPQIKQVPQELNSP.DPPKNEL"
BASE COUNT      734 a    393 c    911 g    744 t
ORIGIN


alignment_scores:
            Quality:   23.00             Length:     23
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x HSU15173    .

 Align seq 1/1  to: HSU15173  from: 1  to: 2382

     257 IleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHisPr 273
         ||||||||||||||||||||||||| || |||||||||||||||||||||
     908 ATTGATAGAAGGTTACGGAAAAATCTAAAATCCCTAATCATTGTACATCC 957

     273 oSerTrpPheIleArgThr 279
         |||||||||||||||||||
     958 TTCTTGGTTTATCAGAACA 976

seq_name: gb_ro:AF035207

seq_documentation_block:
LOCUS       AF035207     2213 bp    mRNA            ROD       20-FEB-1998
DEFINITION  Mus musculus Nip21 mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia, Sciurognathi, Muridae, Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip21 gene [illegible] sequence [illegible] expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, Incyte Pharmaceuticals,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1  .2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit59 and D9Mit1"
     gene            1  .2213
                     /gene="Nip21"
     CDS             134. .1114
                     /gene="Nip21"
                     /note="associates with motifs of both the adenovirus E1B
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP21"
                     /protein_id="AAC01029.1"
                     /db_xref="GI:2911348"
                     /translation="MEGVELKREWQDEDFFIPLPEDDSIEADILAIIGPEDQPGSLEV
                     NGNKVRKKLMAPDISLILDPGDGSLWSDDLDESGEIDLDGLDTPSENSNEFEWEDDLP
                     KPKTTEVIPKGSITFYTAAPPKEDGPPWPMPPIGPQDHPVDMKAIEPYKKVISHGGYY
                     GDGLNAIVVFAVCFMPESSQFNYRYLMDNLFKYVIGTLELLVAENYMIVYLNGAITRR
                     KMPSLGWLPKCYQQIDPRIPKNLKSLIIVHPSWFIPTLLAVTPPFISSKFSQKIPYVF
                     NLAELAELVPMEYVQIPPQIKQVPEELNSPKDPPKNEL"
BASE COUNT      637 a    400 c    971 g    605 t
ORIGIN


alignment_scores:
            Quality:   23.00             Length:     23
```

```
            Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AF035207  ..

 Align seg 1/1  to: AF035207  from: 1  to: 2212

     257 IleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHisPr 273
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     830 ATTGATAGACGATTAACGAAAAACCTTAAGTCTCTAATCATTGTGCACCC 879

     273 oSerTrpPheIleArgThr 279
         |||||||||||||||||||
     880 CTCTTGGTTTATCCGAACA 898

seq_name: gb_htg6.AC009556

seq_documentation_block:
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia.
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,P., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 136385 bases at least Q40
                Consensus quality: 156543 bases at least Q30
                Consensus quality: 165534 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in 120 bases; alarose-fp
                Quality coverage: 4.1 in 120 bases.
            * NOTE: This is a 'working draft' sequence.  It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary.  Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1016: contig of 1016 bp in length
            *                     gap of unknown length
            *     1017      2036: contig of 1020 bp in length
            *                     gap of unknown length
            *     2037      3371: contig of 1335 bp in length
            *                     gap of unknown length
            *     3372      4972: contig of 1601 bp in length
            *                     gap of unknown length
            *     4973      5981: contig of 1009 bp in length
            *                     gap of unknown length
            *     5982      8008: contig of 2027 bp in length
            *                     gap of unknown length
            *     8009      9286: contig of 1278 bp in length
            *                     gap of unknown length
            *     9287     11243: contig of 1957 bp in length
            *                     gap of unknown length
            *    11244     14312: contig of 3069 bp in length
            *                     gap of unknown length
            *    14313     18423: contig of 4111 bp in length
            *                     gap of unknown length
            *    18424     23065: contig of 4642 bp in length
            *                     gap of unknown length
            *    23066     30558: contig of 7493 bp in length
            *                     gap of unknown length
            *    30559     37537: contig of 6979 bp in length
            *                     gap of unknown length
            *    37538     44831: contig of 7294 bp in length
            *                     gap of unknown length
            *    44832     50944: contig of 6113 bp in length
            *                     gap of unknown length
            *    50945     58263: contig of 7319 bp in length
            *                     gap of unknown length
            *    58264     65216: contig of 6953 bp in length
            *                     gap of unknown length
            *    65217     72310: contig of 7094 bp in length
            *                     gap of unknown length
            *    72311     78799: contig of 6489 bp in length
            *                     gap of unknown length
            *    78800     87607: contig of 8808 bp in length
            *                     gap of unknown length
            *    87608     98337: contig of 10730 bp in length
            *                     gap of unknown length
            *    98338    112198: contig of 13861 bp in length
            *                     gap of unknown length
            *   112199    126330: contig of 14132 bp in length
            *                     gap of unknown length
            *   126331    140617: contig of 14287 bp in length
            *                     gap of unknown length
            *   140618    153843: contig of 13226 bp in length
            *                     gap of unknown length
            *   153844    169798: contig of 15955 bp in length
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  44065 t        2 others
ORIGIN
```

```
alignment_scores:
            Quality:   21.00              Length:     21
              Ratio:  1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC009556/rev   ..

 Align seg 1/1  to: reverse of  AC009556  from  1  to  169798

     226 ValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgAr 242
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   13134 GTAGCTGAAGACTATATGATTGTGTACTTGAATGGTGCAACCCCAAGAAG 13085

     242 gArgMetProGly 246
         ||||||||||||
   13084 GAGGATGCCAGGG 13072

seq_name: gb_pri:AB002365

seq_documentation_block:
LOCUS       AB002365     5654 bp    mRNA            PRI       13-FEB-1999
DEFINITION  Human mRNA for KIAA0367 gene, partial cds.
ACCESSION   AB002365
VERSION     AB002365.1  GI:2224674
KEYWORDS    KIAA0367.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
            plus clone:HH0141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1, 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA0367"
     CDS             <1..2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA20822.1"
                     /db_xref="GI:2224675"
                     /translation="ALPMGDPPLSIENPALVPDALLASDICLDISEAAFDHSFSDASG
                     LNTSTGTIDDMSKLILSEGHPEIPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDLL
                     REPEHFLYGGDPPLEEDSLKQSLAPYTPPFDLSYLTEPAQSAETIEEAGSPEDESLGC
                     RAAEIVLSALPDPRSEGNQAFTKNFLPGSQIAVLHIPEDPESVYLPVGAGSNILSPSN
                     VDWEVEIDNSDLPAGGDIGPPNGASKEIPELEEEKTIPTKEPEQIKSEYKEEPCTEKN
                     EDRHALHMDYILVNPEENSHSKPETCEEPESIAELELYVGSKETGLQGTQLASFPDTC
                     QPASLNERKGLSAFKMSSFSIFSSFESPAQDQSWMFLGHSEVGLPSLDARDSGPGWS
                     GKIVEPFSELGLGEGPQLQILEEMKPLESLALEEASGPVSQSQKSKSPGPAGPDAVTH
                     DSEWEMLSPQPVQKNMISDIEMEEETEFLELGTRISRPNGLLSEDVGMDIPFEEGVLS
```

```
                     PSAADMPFEPPNSLDLNETHDRRIKLIAPNINLPLNLSDMILSLNTLAPDEIDINV
                     DELDTPDEADSPEVTPHQPTANKDSGQESESIPEYTAEEDAPLCNSIWPIVVIGEQEQR
                     IDMKVIEPYPPVISHSCVYSDGLNAIIVFAACFLPLGFFMVFVMSYMLLYVISTLE
                     LMVAEDYMIVYLNGATPKKEDARARLDEEMLPDD"
BASE COUNT     1664 a   1188 c   1198 g   1604 t
ORIGIN


alignment_scores:
            Quality    17.00              Length:     17
              Ratio:  1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AB002365   ..

 Align seg 1/1  to: AB002365  from: 1  to: 5654

     187 GlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSerSe 2
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1936 GGTCTAAATGCCATCATTGTGTTTGCTGCTTGTTTTCTGCCAGACAGTAG 1985

     203 r 203
         |
    1986 T 1986

seq_name: gb_pl1:SCN200XG

seq_documentation_block:
LOCUS       SCN200XG     8887 bp    DNA             PLN       6-JUN-1997
DEFINITION  S.cerevisiae N2005, N2007, N2009, N2012, N2014, N2016 genes.
ACCESSION   X77114
VERSION     X77114.1  GI:496710
KEYWORDS    mitochondrial protein; MRP7 protein.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 8887)
  AUTHORS   Verhasselt,P., Aert,R., Volckaert,G. and Volckaert,G.
  TITLE     Nucleotide sequence analysis of an 8887 bp region of the left arm
            of yeast chromosome XIV encompassing the centromere sequence
  JOURNAL   Yeast 10 (7), 945-951 (1994)
  MEDLINE   95076713
REFERENCE   2  (bases 1 to 8887)
  AUTHORS   Volckaert,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1994) G. Volckaert, Laboratory of Gene
            Technology, Catholic University Leuven, W. Decroylaan 42, 3001
            Leuven, BELGIUM
REFERENCE   3  (bases 1 to 8887)
  AUTHORS   Kemp,J.T., Balasubramanian M.K. and Gould K.L.
  TITLE     A wat1 mutant of fission yeast is defective in cell morphology
  JOURNAL   Mol. Gen. Genet. 254 (2), 127-138 (1997)
  MEDLINE   97261860
FEATURES             Location/Qualifiers
     source          1..8887
                     /organism="Saccharomyces cerevisiae"
                     /strain="FY1679"
                     /db_xref="taxon:4932"
                     /chromosome="14"
                     /clone_lib="B. Dujon"
                     /clone="cosmid 14-20"
                     /map="centromere region"
     misc_feature    complement(1..264)
                     /note="centromere"
     gene            1..8887
                     /gene="DOM34"
                     /db_xref="SGD:S0004946"
     gene            1..8887
                     /gene="MRP7"
                     /db_xref="SGD:S0004950"
     gene            1..8887
```

```
                /gene="RLP7"
                /db_xref="SGD:S0004947"
CDS             complement(290..1401)
                /note="ORF N2016"
                /codon_start=1
                /protein_id="CAA54377.1"
                /db_xref="GI:496711"
                /db_xref="SWISS-PROT:P33309"
                /translation="MKVISLKKDSFNKGGAVITILPEDKEDLFTVYYQIVDKDDELIFK
                KKFTSKLDEAGKKKSTDLVKIKIKVISEDFDMKDEYLKYKGVTVTDESGASNVDIPVG
                KYLSFTLDYVYPFTIIKQNFNKFMQKLLNEACNIEYKSDTAAVVLQEGIAHVCLVISS
                STILKQKIEYSMPKKKRTIDVLKFDEKTEKFYKAIYSAMKKDLNFDKLKTIILCSPGF
                YAKILMDKIFQYAEEEHNKKILDNKGMFFIAHCSTGYLQGINEVLKNPLYASKLQDTK
                YSKEIMVMDEFLLHLNKDDLKAWYGEKEVVKAAEYGAISYLLLTDKVLHSDNIAQREE
                YLKLMDSVESNGGKALVLSTLHSLGEELDQLTGIACILKYPLPDLDEDDGEE"
CDS             1764..2732
                /gene="RLP7"
                /note="ORF N2014"
                /codon_start=1
                /db_xref="SGD:S0004947"
                /protein_id="CAA54376.1"
                /db_xref="GI:496712"
                /db_xref="SWISS-PROT:P40693"
                /translation="MSSTQLSKAQTLNSNPEILLRKRRNADPTPIEPQELAKKKREEI
                IKKKPSNKNKFVPAESIVAKTLATSPEKEPIKPVSILEDKKAKNETQHIASGKDFILK
                ITEKANGAEENSVDLEETEEEEDDGLIPEKTTYDGKPALLFIVRVRGPLAVNIPNKAF
                KILSLLRLVETNTGVFVKLTKNVYPLLKVIAPYVVIGKPSLSSIPSLIQKPGRIIYKG
                ENEAPHEIVLNDNNIVEEQLSDHGIICVEDIIHEIAIMGESFSVCNFFLQPFKLNRE
                VSGFGSLNRLRKIKQREAESRIRQFSNAATAPVIEVDIDSLLAKLN"
CDS             3077..3931
                /gene="RLP7"
                /note="ORF N2012"
                /codon_start=1
                /db_xref="SGD:S0004947"
                /protein_id="CAA54377.1"
                /db_xref="GI:496713"
                /db_xref="SWISS-PROT:P38921"
                /translation="MNIFFLSLLSGAAAGTSTDLVFFPIDTIKTRLQAKGGFFANGGY
                KGIYRGLGSAVVASAPGASLFFISYDYMKVKSRPYISKLYSQGSEQLIDTTTHMLSSS
                IGEICACLVRVPAEVVKQRTQVHSTNSSWQTLQSILRNDNKEGLRKNLYRGWSTTIMR
                EIPFTCIQFPLYEYLKKTWAKANGQSQVEPWKGAICGSIAGGIAAAITTPLDFLKTRL
                MLNKTTASLGSVIIRIYREEGPAVFFSGVGPRTMWISAGGAIFLGMYETVHSLLSKSF
                PTAGEMRA"
CDS             complement(4285..5574)
                /note="N2009"
                /codon_start=1
                /protein_id="CAA54378.1"
                /db_xref="GI:496714"
                /db_xref="SWISS-PROT:P38922"
                /translation="MSDDHGYERDNHLSRRSGNYNGRRKFADTYRGSRDRGEYRGGRE
                RSDYRERERFNNRDNPRSRDRYDDRRRGRDVTGRYGNRRDDYPRSFRSRHNTRDDSRR
                GGFGSSGARGDYGPLLARELDSTYEEKVNRNYSNSIFVGNLTYDSTPEDLIEFFSQIG
                KVVRADIITSRGHHRGMGTVEFTNSDDVDRAIRQYDGAFFMDRKIFVRQDNPPPSNNI
                KERKALDRGELRHNRKIHEVIVKNLPASVNWQALKDIFKECGNVAHADVELDGDGVST
                GSGTVSFYDIKDLHRAIEKYNGYSIEGNVLDVKSKESVHNHSDGDDVDIPMDDSPVNE
                EARKFTENVVGGGERNRLIYCSNLPFSTAKSDLYDLFETIGKVNNAELRYDSKGAPTG
                IAVVEYDNVDDADVCIERLNNYNYGGCDLDISYAKRL"
CDS             6478..7593
                /gene="RLP7"
                /note="ORF N2007"
                /codon_start=1
                /db_xref="SGD:S0004947"
                /protein_id="CAA54379.1"
                /db_xref="GI:496715"
                /db_xref="SWISS-PROT:P12687"
                /translation="MWNRILLQTSSFSFQKHVSGVFLQVRNATKRAAGSRISMKDSAG
                RRLGPKKYEGQDVSTGEIIMRQRGTKFYPGENVGIGKDHSIFALEPGVVRYYLDPFHP
                KRKFIGVALRRDLKLPSPHFEPTVRRFGRFELTNKRAAYKEENSISPKDYLAKPNILK
                QLEVRESKRKELQDKLSKVLRDELKLDIKDIELATSYLIPVRASLKNGYPIEDARFNS
                RYYLKEEERLKAREESWTNEKLSESLSKIEDTSDLLNSSTSFNNKLELHQYISEIEKI
                ALKAKLLEDLEKSQHLETKKDKNYIKALFKDACNFLTLSEEVHLRRKYLKSVFPETDS
                TVETKSGKKSIVSPRLDYTKNKVEVIAPSRPAFLSKL"
CDS             complement(7928..8839)
```

```
                /note="ORF N..."
                /codon_start=1
                /protein_id="CAA54..."
                /db_xref="GI:..."
                /db_xref="SWISS-PROT:P41..."
                /translation="...UNFLEITNIK
                KLLAIACHLNVL...ELDCTIKVW
                DVRSFSIIRN...HQLIPEDDT
                SLQSLSMASD...ITRILLSSD
                VKHLAICSADH...NTASSDHYVRL
                WDLSIRLIVRQL...AVIVALNDV"
BASE COUNT     2749 a   1672 c   1741 g   2825 t
ORIGIN

alignment_scores:
          Quality:    11.00            Length:
            Ratio:    1.000              Gaps:
 Percent Similarity: 100.000   Percent Identity:

alignment_block:
 US-09-174-937-2 x SCN200XG ..

 Align seq 1/1  to: SCN200XG  from: 1  to: 8847

    324 GluGluGluArgLeuLysAlaArgAr...Ser 334
        | | | | | | | | | | | | | | |
   7147 GAAGAAGAGCGTTTGAAGGCAAG...AAG 7179

seq_name. gb_pl1.SCYNL005C

seq_documentation_block:
LOCUS       SCYNL005C    2353 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome XIV reading frame ORF YNL005c.
ACCESSION   Z71281 Y13139
VERSION     Z71281.1  GI:1301819
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Dolgnon,F. and Crouzet,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2353)
  AUTHORS   Aert,R., Verhasselt,P., Voet,M. and Volckaert,G.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2353)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XIV sequencing project, MIPS at the
            Max-Planck-Institut f. Biochemie, Am Klopferspitz 18a, D-82152
            Martinsried, FRG. E-mail: ...
FEATURES             Location/Qualifiers
     source          1..2353
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="XIV"
     gene            complement(...1245)
                     /gene="MRPL2"
                     /db_xref="SGD:S...495..."
     CDS             complement(...1245)
                     /gene="MRPL2"
                     /note="ORF YNL005c"
                     /codon_start=1
                     /db_xref="SGD:S...4951"
                     /protein_id="CAA...64.1"
                     /db_xref="GI:...1820"
                     /db_xref="SWISS-PROT:...12687"
                     /translation="MWN...RISMKDSAG
                     RRLGPKKYEGQ...VVRYYLDPFHP
                     KRKFIGVALRR...YLAKPNILK
```

```
                     QLEVPESKPKPEIQDKLSKVLKLDELKLDIKDIELATSYLIPVPASLKNQYPIEDARFNS
                     PYYLKEPEPLKAPPESWTNEPLSESLSKIDFDSDLLNSSTSFNNKLELHQYISEQEPQ
                     ALKAKLLEDLEKSQHLETKKDKNYIKALPKDACNELTLSEEVHLPPKYLKSVFPETDS
                     TVETKSGKKSIVSPPEDYTKNFVEVIAPSPPAFLSKL"
BASE COUNT      717 a    456 c    397 g    783 t
ORIGIN


alignment_scores:
            Quality:   11.00             Length:     11
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x SCYNI005C/rev

 Align seq 1/1  to: reverse of  SCYNI005C  from  1  to: 2353

     324 GluGluGluArgLeuLysAlaArgArgGluSer 334
         ||||||||||||||||||||||||||||||||||
     781 GAAGAAGAGCGTTTGAAGGCAAGGAGAGAAAGT 749

seq_name: gb_pl1:YSCMRP7

seq_documentation_block:
LOCUS       YSCMRP7      2016 bp    DNA             PLN       27-APR-1993
DEFINITION  Yeast (S cerevisiae) mitochondrial 40 kd ribosomal protein large
            subunit gene, complete cds.
ACCESSION   M22116
VERSION     M22116.1  GI:171983
KEYWORDS    ribosomal protein.
SOURCE      Yeast (S cerevisiae; strain S288C) DNA, (library of M.Rose), clone
            lambda-MRP7.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2016)
  AUTHORS   Fearon,K. and Mason,T.L.
  TITLE     Structure and regulation of a nuclear gene in Saccharomyces
            cerevisiae that specifies MRP7, a protein of the large subunit of
            the mitochondrial ribosome
  JOURNAL   Mol. Cell. Biol. 8, 3636-3646 (1988)
  MEDLINE   89127202
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by K.Fearon, 01-MAR-1989.
FEATURES             Location/Qualifiers
     source          1..2016
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     gene            1..2016
                     /gene="MRP7"
                     /db_xref="SGD:S0004950"
     mRNA            352..>1536
                     /gene="MRP7"
                     /note="RPLS mRNA"
                     /db_xref="SGD:S0004951"
     mRNA            378..>1536
                     /gene="MRP7"
                     /note="RPLS mRNA"
                     /db_xref="SGD:S0004951"
     sig_peptide     421..501
                     /gene="MRP7"
                     /note="40 kd ribosomal protein large subunit signal
                     peptide (put.); putative"
                     /db_xref="SGD:S0004951"
     CDS             421..1536
                     /gene="MRP7"
                     /note="40 kd ribosomal protein large subunit precursor"
                     /codon_start=1
                     /db_xref="SGD:S0004950"
                     /protein_id="AAA34791.1"
                     /db_xref="GI:171984"
                     /translation="MWNRLLTTSRFSPKHVSRVFIQVRNATPAAASSPTSMKESAT
                     PPLGPKKYEDQIVSTDLIMPLPGTKFYMTPNVTLKHSIPALDTVVRYYLDPFHP
                     KPKFIGVALPRTPLESPHFETVPSPGPFLTNPRAAPPFPNTPAPLALKPNILK
                     QLEVPESKPKPEIQDKLSKVLKLDELKLDIKDIELATSYLIPVPASLKNQYPIEDARFNS
                     PYYLKEPEPLKAPPESWTNEPLSESLSKIDFDSDLLNSSTSFNNKLELHQYISEQEPQ
                     ALKAKLLEDLEKSQHLETKKDKNYIKALPKDACNELTLSEEVHLPPKYLKSVFPETDS
                     TVETKSGKKSIVSPPEDYTKNFVEVIAPSPPAFLSKL"
     mat_peptide     502..1533
                     /gene="MRP7"
                     /note="40 kd ribosomal protein large subunit"
                     /db_xref="SGD:S0004950"
BASE COUNT      663 a    338 c    409 g    606 t
ORIGIN      399 bp upstream of HindIII site.


alignment_scores
            Quality:   11.00             Length:     11
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x YSCMRP7    ..

 Align seq 1/1  to: YSCMRP7  from: 1  to: 2016

     324 GluGluGluArgLeuLysAlaArgArgGluSer 334
         ||||||||||||||||||||||||||||||||||
    1090 GAAGAAGAGCGTTTGAAGGCAAGGAGAGAAAGC 1123

seq_name: gb_vi:AF123486

seq_documentation_block:
LOCUS       AF123486     4513 bp    mRNA            VRL       15-SEP-1999
DEFINITION  San Angelo virus polyprotein mRNA, complete cds.
ACCESSION   AF123486
VERSION     AF123486.1  GI:5823116
KEYWORDS    .
SOURCE      San Angelo virus.
  ORGANISM  San Angelo virus
            Viruses; ssRNA negative-strand viruses; Bunyaviridae; Bunyavirus;
            Z-California encephalitis virus Group.
REFERENCE   1  (bases 1 to 4513)
  AUTHORS   Campbell,W.P. and Huang,C.
  TITLE     Sequence comparisons of medium RNA segment among [illegible]nia
            serogroup viruses
  JOURNAL   Virus Res. 61 (2), 137-144 (1999)
  MEDLINE   99402068
REFERENCE   2  (bases 1 to 4513)
  AUTHORS   Campbell,W. and Huang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1999) Wadsworth Center, New York State Department
            of Health, Box 509, Albany, NY 12201-0509, USA
FEATURES             Location/Qualifiers
     source          1..4513
                     /organism="San Angelo virus"
                     /strain="ATCC VR-723"
                     /db_xref="taxon:[illegible]"
                     /db_xref="ATCC:VR-723"
                     /chromosome="M segment"
     CDS             61..4386
                     /note="cleaved into glycoproteins G1 and G2"
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAD53042.1"
                     /db_xref="GI:5823117"
                     /translation="MIFLVTLVSAGIVFLRTQLSAIVPLN[illegible]
                     VSMIKTEARYVPNSSGVFSNNVALPKWLVSDWHDVPPFVSIGP[illegible]
                     ESYVCSADCIIGIIEIAPLIPLTIFINNHFVAGIAYLWKSINPQTCEHLK
                     VSCGPKSIQFHACFNGHMSCVRFLHRTIPRSMANSICQNIEIIILLIILALLIFLLLT
                     ILSKTYICYLLNPIIIAHLIGLIYNKSCKKCLTPYVVHRPTAPTHVCGARYET
                     SDPMKLHFASGIVVVASEAAPVMCPAPNAN[illegible]PVTLNGLTIGDS
                     KEVLNPDELFDKMLLHAPLNMYCLICINYIAIWLIIILAVSIIFKYQHRFYNY
                     YAMYCLECEMYHKSLSWNGLPTNKCSLCIDSVLLAMYLSHPIDTLLYKVVW
                     VMNFLTIYVALFVSFMIVIAATTPPVTPTPTVNNTTVVVNTCEPKDWKE
```

```
                     SYSNIAEQIKGLQAISVLDIPLIAPIPEDISGGLRYIESIPTHHAQLTAEYAMLIPYC
                     DYHKLFDFNSDYSQIIWPMYIPSHDFDACILYPNQHFDPCVKHGEKDSSSSWDFASEM
                     KNYYSGKQGKFNKDLNLALQSIHHAFPGTSSSYIATILSEKPNDQLVEYMEKLPEKFI
                     GNALLPAVLDFITYMKSIIPMTNFKHDEFWDEFVHLPAPTKAPSMPPSGNENYDFYAP
                     TSPAGTKSCKNIKNVVCLSPPSGASYEGIVACGEPSAPQIFKMPSESIYHNNADQSMY
                     CIADTHCLQDYSPVEDEELNAIKKSKCWETEFTNAVPFKQSDGIPSCPMPDSGNCNVI
                     INLWPVMQCDDGKFYYSEHCKYDKDQDIGHFCLSPPCNMVPFPISEKNILACDWIVS
                     HPNIDKIDVHSLEDLEQYKKAITQKLQTSLSIFKYAKTMNLPHVKPIYKYISIEGTET
                     AEGVENAFIESEIPALAGTSIGFKVNSKDGKHIMDIIGYVKTASYSSVYTKLYTISPI
                     IGVNTKHDELCTGPCPKQIEHKNGWITFSKERTSSWGCEEFGCLAISDGCVFGSCQDI
                     IKDEMSVYPKETEETTIVELCLIFSDKTYCVELNPVTPIITDLFEVQFKTVESYTLPR
                     IVAVKNHEVMVGQINDLGVYSKGCGNVQKVNGTVYGNGVPKFDYLCHLASPKEVIIPY
                     CFDNDYQACKFLQSPSSYPLVEDASTVSIIDYKKILGTIKMKAILGDVPYKTFSESID
                     ISAEGTCTGCVNCFENIHCEFTIHSTVEASCPVVSTCTLFHDPILIGPDNHKYAMKVI
                     CTEAPKTSLSFKICNTKVDAAITIVDSKPIVELAPVDQTAYIPEKDDPCKIWMCPVPD
                     EGMQVLSEPFKNLFGSYIGIFYTGLAVLIIALIAVYIVMPICFKVKDILKKHDDAYRR
                     ELKIR"
     mat_peptide     61..957
                     /evidence=not_experimental
                     /product="glycoprotein G2"
     mat_peptide     1480..4383
                     /evidence=not_experimental
                     /product="glycorprotein G1"
BASE COUNT     1557 a    771 c    911 g   1274 t
ORIGIN


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AF123496   ..

 Align seq 1/1  to: AF123486  from: 1  to: 4513

     299 ValHisSerLeuGluAspLeuGlnGln 307
         ||||||||||||||||||||||||||||
    2927 GTCCATTCTCTTGAAGATCTACAGCAA 2955

seq_name: gb_htg3:AC009344

seq_documentation_block:
LOCUS       AC009344   113791 bp    DNA             HTG       13-SEP-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR10N14 (BL961)
            RPCI-98 20.N.14 map 93A-93B strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 93 unordered pieces.
ACCESSION   AC009344
VERSION     AC009344.4  GI:5970954
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 113791)
  AUTHORS   Celniker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Gallo,P., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weiberg,T., Zhang,P., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113791)
  AUTHORS   Celniker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Gallo,P., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weiberg,T., Zhang,P., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Theoretical Biology, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 13, 1999 this sequence version replaced gi:5970954.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to help@fruitfly.berkeley.edu. All numbers in this submission meet
            the following cutoffs: length >= 20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 93 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the annotation from it will
            * be preserved.
            *        1      724: contig of 724 bp in length
            *      725      804: gap of unknown length
            *      805     1370: contig of 566 bp in length
            *     1371     1450: gap of unknown length
            *     1451     2301: contig of 851 bp in length
            *     2302     2381: gap of unknown length
            *     2382     3626: contig of 1245 bp in length
            *     3627     3756: gap of unknown length
            *     3757     4723: contig of 967 bp in length
            *     4724     4803: gap of unknown length
            *     4804     5459: contig of 656 bp in length
            *     5460     5539: gap of unknown length
            *     5540     6558: contig of 1019 bp in length
            *     6559     6638: gap of unknown length
            *     6639     7683: contig of 1045 bp in length
            *     7684     7763: gap of unknown length
            *     7764     8432: contig of 669 bp in length
            *     8433     8562: gap of unknown length
            *     8563     9140: contig of 578 bp in length
            *     9141     9220: gap of unknown length
            *     9221     9858: contig of 638 bp in length
            *     9859     9939: gap of unknown length
            *     9940    10609: contig of 670 bp in length
            *    10610    10695: gap of unknown length
            *    10696    11257: contig of 562 bp in length
            *    11258    11338: gap of unknown length
            *    11339    12413: contig of 1076 bp in length
            *    12414    12493: gap of unknown length
            *    12494    13470: contig of 977 bp in length
            *    13471    13550: gap of unknown length
            *    13551    14402: contig of 852 bp in length
            *    14403    14482: gap of unknown length
            *    14483    15369: contig of 887 bp in length
            *    15370    15449: gap of unknown length
            *    15450    16379: contig of 930 bp in length
            *    16380    16459: gap of unknown length
            *    16460    17477: contig of 1018 bp in length
            *    17478    17557: gap of unknown length
            *    17558    18729: contig of 1172 bp in length
            *    18730    18789: gap of unknown length
            *    18790    19810: contig of 1021 bp in length
            *    19811    19890: gap of unknown length
            *    19891    20622: contig of 732 bp in length
            *    20623    20702: gap of unknown length
            *    20703    21658: contig of 956 bp in length
            *    21659    21738: gap of unknown length
            *    21739    22507: contig of 769 bp in length
            *    22508    22587: gap of unknown length
            *    22588    23621: contig of 1034 bp in length
            *    23622    23702: gap of unknown length
            *    23703    24755: contig of 1053 bp in length
```

```
*    24756    24835: gap of unknown length
*    24836    25985: contig of 1150 bp in length
*    25986    26065: gap of unknown length
*    26066    26773: contig of 708 bp in length
*    26774    26853: gap of unknown length
*    26854    27859: contig of 1006 bp in length
*    27860    27939: gap of unknown length
*    27940    29056: contig of 1117 bp in length
*    29057    29136: gap of unknown length
*    29137    30706: contig of 1570 bp in length
*    30707    30786: gap of unknown length
*    30787    31969: contig of 1183 bp in length
*    31970    32049: gap of unknown length
*    32050    33506: contig of 1457 bp in length
*    33507    33586: gap of unknown length
*    33587    34873: contig of 1287 bp in length
*    34874    34953: gap of unknown length
*    34954    35703: contig of 750 bp in length
*    35704    35783: gap of unknown length
*    35784    36849: contig of 1066 bp in length
*    36850    36929: gap of unknown length
*    36930    38514: contig of 1585 bp in length
*    38515    38594: gap of unknown length
*    38595    39790: contig of 1196 bp in length
*    39791    39870: gap of unknown length
*    39871    41219: contig of 1349 bp in length
*    41220    41299: gap of unknown length
*    41300    42739: contig of 1440 bp in length
*    42740    42819: gap of unknown length
*    42820    43187: contig of 368 bp in length
*    43188    43267: gap of unknown length
*    43268    44951: contig of 1684 bp in length
*    44952    45031: gap of unknown length
*    45032    46136: contig of 1105 bp in length
*    46137    46216: gap of unknown length
*    46217    47443: contig of 1227 bp in length
*    47444    47523: gap of unknown length
*    47524    49450: contig of 1927 bp in length
*    49451    49530: gap of unknown length
*    49531    51959: contig of 2428 bp in length
*    51960    52038: gap of unknown length
*    52039    53454: contig of 1416 bp in length
*    53455    53534: gap of unknown length
*    53535    54942: contig of 1408 bp in length
*    54943    55022: gap of unknown length
*    55023    57299: contig of 2276 bp in length
*    57300    57378: gap of unknown length
*    57379    58898: contig of 1520 bp in length
*    58899    58978: gap of unknown length
*    58979    60318: contig of 1340 bp in length
*    60319    60398: gap of unknown length
*    60399    62444: contig of 2046 bp in length
*    62445    62524: gap of unknown length
*    62525    64809: contig of 2285 bp in length
*    64810    64889: gap of unknown length
*    64890    66814: contig of 1925 bp in length
*    66815    66894: gap of unknown length
*    66895    69113: contig of 2219 bp in length
*    69114    69193: gap of unknown length
*    69194    70960: contig of 1767 bp in length
*    70961    71040: gap of unknown length
*    71041    74483: contig of 3443 bp in length
*    74484    74563: gap of unknown length
*    74564    76539: contig of 1976 bp in length
*    76540    76619: gap of unknown length
*    76620    78356: contig of 1737 bp in length
*    78357    78436: gap of unknown length
*    78437    80218: contig of 1782 bp in length
*    80219    80298: gap of unknown length
*    80299    82007: contig of 1709 bp in length
*    82008    82087: gap of unknown length
*    82088    85187: contig of 3100 bp in length
*    85188    85267: gap of unknown length
*    85268    89464: contig of 4197 bp in length
*    89465    89544: gap of unknown length
*    89545    93275: contig of 3731 bp in length
*    93276    93355: gap of unknown length
*    93356    93955: contig of 600 bp in length
*    93956    94035: gap of unknown length
*    94036    94599: contig of 564 bp in length
*    94600    94679: gap of unknown length
*    94680    95159: contig of 480 bp in length
*    95160    95239: gap of unknown length
*    95240    95838: contig of 599 bp in length
*    95839    95918: gap of unknown length
*    95919    96620: contig of 702 bp in length
*    96621    96700: gap of unknown length
*    96701    97360: contig of 660 bp in length
*    97361    97440: gap of unknown length
*    97441    98053: contig of 613 bp in length
*    98054    98133: gap of unknown length
*    98134    98730: contig of 597 bp in length
*    98731    98810: gap of unknown length
*    98811    99454: contig of 644 bp in length
*    99455    99534: gap of unknown length
*    99535   100117: contig of 583 bp in length
*   100118   100197: gap of unknown length
*   100198   100805: contig of 608 bp in length

alignment_scores:
            Quality:    9.00              Length:       9
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC009344/rev   ..

 Align seq 1/1  to reverse of: AC009344  from: 1  to: 114741

     267 SerLeuIleIleValHisProSerTrp 275
         ||||||||||||||||||||||||||||
  111936 AGTTTAATAATCGTTCATCCAAGCTGG 111910

seq_name: gb_htg3:AC009348

seq_documentation_block.
LOCUS       AC009348   139296 bp    DNA             HTG        8-SEP-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR05I01 (BACR) RPCI-98
            05 I 1 map 93A-93B strain y cn bw sp  *** SEQUENCING IN PROGRESS
            ***, 117 unordered pieces.
ACCESSION   AC009348
VERSION     AC009348.3  GI:5836098
KEYWORDS    HTG; HTGS_PHASE1
SOURCE      fruit fly
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota, Neoptera, Endopterygota, Diptera, Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 139296)
  AUTHORS   Celniker,S.E., Agbayani,A., Akiyama,T.T., Baxter,E., Black,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Chiswinska,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.F., Kerr,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Parker,M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,I., Zhang,M., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139296)
  AUTHORS   Celniker,S.E., Agbayani,A., Akiyama,T.T., Baxter,E., Black,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Chiswinska,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.F., Kerr,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
```

Moshrefi,A P , Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B , Poon,I , Sequeira,A., Sethi,H., Snir,E.,
Swirskas,P P., Wan,K H., Weinburg,T., Zhang,R., Zideran,D.L. and
Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Sep 8, 1999 this sequence version replaced gi:5832514.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly berkeley edu All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 117 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1      530: contig of 530 bp in length
*      531      610: gap of unknown length
*      611     1028: contig of 418 bp in length
*     1029     1108: gap of unknown length
*     1109     1689: contig of 581 bp in length
*     1690     1769: gap of unknown length
*     1770     2360: contig of 591 bp in length
*     2361     2440: gap of unknown length
*     2441     2875: contig of 435 bp in length
*     2876     2955: gap of unknown length
*     2956     3507: contig of 552 bp in length
*     3508     3587: gap of unknown length
*     3588     4170: contig of 583 bp in length
*     4171     4250: gap of unknown length
*     4251     4851: contig of 601 bp in length
*     4852     4931: gap of unknown length
*     4932     5407: contig of 476 bp in length
*     5408     5487: gap of unknown length
*     5488     6192: contig of 705 bp in length
*     6193     6272: gap of unknown length
*     6273     7062: contig of 790 bp in length
*     7063     7142: gap of unknown length
*     7143     7923: contig of 781 bp in length
*     7924     8003: gap of unknown length
*     8004     8706: contig of 703 bp in length
*     8707     8786: gap of unknown length
*     8787     9535: contig of 749 bp in length
*     9536     9615: gap of unknown length
*     9616    10265: contig of 650 bp in length
*    10266    10345: gap of unknown length
*    10346    10963: contig of 618 bp in length
*    10964    11043: gap of unknown length
*    11044    11782: contig of 739 bp in length
*    11783    11862: gap of unknown length
*    11863    12493: contig of 631 bp in length
*    12494    12573: gap of unknown length
*    12574    13467: contig of 894 bp in length
*    13468    13547: gap of unknown length
*    13548    14166: contig of 619 bp in length
*    14167    14246: gap of unknown length
*    14247    14852: contig of 606 bp in length
*    14853    14932: gap of unknown length
*    14933    15725: contig of 793 bp in length
*    15726    15805: gap of unknown length
*    15806    16466: contig of 661 bp in length
*    16467    16546: gap of unknown length
*    16547    17518: contig of 972 bp in length
*    17519    17598: gap of unknown length
*    17599    18426: contig of 828 bp in length
*    18427    18506: gap of unknown length
*    18507    19208: contig of 702 bp in length
*    19209    19288: gap of unknown length
*    19289    19955: contig of 667 bp in length
*    19956    20035: gap of unknown length
*    20036    20703: contig of 668 bp in length
*    20704    20783: gap of unknown length
*    20784    21357: contig of 574 bp in length
*    21358    21437: gap of unknown length
*    21438    22578: contig of 1141 bp in length
*    22579    22658: gap of unknown length
*    22659    23489: contig of 831 bp in length
*    23490    23569: gap of unknown length
*    23570    24571: contig of 1002 bp in length
*    24572    24651: gap of unknown length
*    24652    25871: contig of 1220 bp in length
*    25872    25951: gap of unknown length
*    25952    27368: contig of 1417 bp in length
*    27369    27448: gap of unknown length
*    27449    28788: contig of 1340 bp in length
*    28789    28868: gap of unknown length
*    28869    29655: contig of 787 bp in length
*    29656    29735: gap of unknown length
*    29736    30970: contig of 1235 bp in length
*    30971    31050: gap of unknown length
*    31051    32208: contig of 1158 bp in length
*    32209    32338: gap of unknown length
*    32339    33452: contig of 1114 bp in length
*    33453    33532: gap of unknown length
*    33533    34609: contig of 1077 bp in length
*    34610    34689: gap of unknown length
*    34690    35782: contig of 1093 bp in length
*    35783    35862: gap of unknown length
*    35863    36900: contig of 1038 bp in length
*    36901    36980: gap of unknown length
*    36981    37638: contig of 658 bp in length
*    37639    37718: gap of unknown length
*    37719    38413: contig of 695 bp in length
*    38414    38493: gap of unknown length
*    38494    39401: contig of 908 bp in length
*    39402    39481: gap of unknown length
*    39482    40198: contig of 717 bp in length
*    40199    40278: gap of unknown length
*    40279    41009: contig of 731 bp in length
*    41010    41089: gap of unknown length
*    41090    42014: contig of 925 bp in length
*    42015    42094: gap of unknown length
*    42095    43956: contig of 1862 bp in length
*    43957    44036: gap of unknown length
*    44037    45014: contig of 978 bp in length
*    45015    45094: gap of unknown length
*    45095    45773: contig of 679 bp in length
*    45774    45853: gap of unknown length
*    45854    46728: contig of 875 bp in length
*    46729    46808: gap of unknown length
*    46809    48098: contig of 1290 bp in length
*    48099    48178: gap of unknown length
*    48179    49077: contig of 899 bp in length
*    49078    49157: gap of unknown length
*    49158    50092: contig of 935 bp in length
*    50093    50172: gap of unknown length
*    50173    51662: contig of 1490 bp in length
*    51663    51742: gap of unknown length
*    51743    53033: contig of 1291 bp in length
*    53034    53413: gap of unknown length
*    53414    55264: contig of 1851 bp in length
*    55265    55344: gap of unknown length
*    55345    55979: contig of 635 bp in length
*    55980    56059: gap of unknown length
*    56060    57051: contig of 992 bp in length
*    57052    57131: gap of unknown length
*    57132    58802: contig of 1671 bp in length
*    58803    58882: gap of unknown length
*    58883    60414: contig of 1532 bp in length
*    60415    60494: gap of unknown length
*    60495    61949: contig of 1455 bp in length
```

```
            *    61950    62029: gap of unknown length
            *    62030    62783: contig of 754 bp in length
            *    62784    62863: gap of unknown length
            *    62864    63951: contig of 1088 bp in length
            *    63952    64031: gap of unknown length
            *    64032    65927: contig of 1896 bp in length
            *    65928    66007: gap of unknown length
            *    66008    67701: contig of 1694 bp in length
            *    67702    67781: gap of unknown length
            *    67782    69314: contig of 1533 bp in length
            *    69315    69394: gap of unknown length
            *    69395    70711: contig of 1317 bp in length
            *    70712    70791: gap of unknown length
            *    70792    72499: contig of 1707 bp in length
            *    72499    72578: gap of unknown length
            *    72579    73750: contig of 1172 bp in length
            *    73751    73830: gap of unknown length
            *    73831    75144: contig of 1314 bp in length
            *    75145    75224: gap of unknown length
            *    75225    76432: contig of 1208 bp in length
            *    76433    76512: gap of unknown length
            *    76513    78452: contig of 1940 bp in length
            *    78453    78532: gap of unknown length
            *    78533    80326: contig of 1794 bp in length

alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC009348   ..

 Align seq 1/1  to: AC009348  from: 1  to: 149296

     267 SerLeuIleIleValHisProSerTrp 275
         |||||||||||||||||||||||||||
  110502 AGTTTAATAATCGTTCATCCAAGCTGG 110528

seq_name: gb_htg3:AC010754

seq_documentation_block:
LOCUS        AC010754   187096 bp    DNA             HTG       22-SEP-1999
DEFINITION   Homo sapiens chromosome 18 clone 403_A_21 map 18, *** SEQUENCING IN
             PROGRESS ***, 10 unordered pieces.
ACCESSION    AC010754
VERSION      AC010754.1  GI:5918009
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 187096)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 18, clone 403_A_21
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 187096)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
             Galagan,J., Gardyna,S., Grant,J., Hages,E., Heaford,A., Horton,L.,
             Howland,J., Johnson,E., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
             McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
             Morrow,J., Naylor,J., Norman,J.H., O'Connor,T., O'Donnell,P.,
             Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
             Wyman,D., Yo,W.J., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      All repeats were identified using RepeatMasker [redacted]
             Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 10 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1     5032: contig of 5032 bp in length
             *                   gap of unknown length
             *     5033     9552: contig of 4520 bp in length
             *                   gap of unknown length
             *     9553    17303: contig of 7751 bp in length
             *                   gap of unknown length
             *    17304    31421: contig of 14118 bp in length
             *                   gap of unknown length
             *    31422    47927: contig of 16506 bp in length
             *                   gap of unknown length
             *    47928    73892: contig of 25965 bp in length
             *                   gap of unknown length
             *    73893   102236: contig of 28344 bp in length
             *                   gap of unknown length
             *   102237   135888: contig of 33652 bp in length
             *                   gap of unknown length
             *   135889   184525: contig of 48637 bp in length
             *                   gap of unknown length
             *   184526   187096: contig of 2571 bp in length.
FEATURES             Location/Qualifiers
     source          1..187096
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="403_A_21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    53592 a   38891 c   39209 g   54207 t    1197 others
ORIGIN


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AC010754   ..

 Align seq 1/1  to: AC010754  from: 1  to: 187096

      44 ThrSerSerProProAsnThrLeuAsn 52
         |||||||||||||||||||||||||||
   62874 ACGTCGTCTCCTCCCAACACCCTAAAT 62900

seq_name: gb_ba2:AF065404

seq_documentation_block:
LOCUS        AF065404   181654 bp    DNA   circular  BCT       [illegible]
DEFINITION   Bacillus anthracis virulence plasmid pX01, complete sequence.
ACCESSION    AF065404
VERSION      AF065404.1  GI:4894216
KEYWORDS     .
SOURCE       Bacillus anthracis.
  ORGANISM   Bacillus anthracis
             Bacteria; Firmicutes; Bacillus/Clostridium group;
             Bacillus/Staphylococcus group; Bacillus.
REFERENCE    1  (bases 1 to 181654)
  AUTHORS    Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A.R., Hill,K.K.,
             Keim,P., Koehler,T.M., Lamke,G., Kumano,S., Mahillon,J., Munter,D.,
             Martinez,Y., Ricke,D., Svensson,R. and Jackson,P.J.
```

```
TITLE       Sequence and organization of pXO1, the large Bacillus anthracis
            plasmid harboring the anthrax toxin genes
  JOURNAL   J. Bacteriol. 181 (20), 6509-6515 (1999)
  MEDLINE   99445483
REFERENCE   2  (bases 1 to 181654)
  AUTHORS   Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A., Hill,K., Keim,P.,
            Koehler,T., Lamke,G., Kumano,S., Manter,D., Martinez,Y.,
            Ricke,D.O., Svensson,R. and Jackson,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) Life Sciences Division, Los Alamos National
            Laboratory, TA43, LS-6, HRL-1, MS M888, Los Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1..181654
                     /organism="Bacillus anthracis"
                     /strain="Sterne"
                     /db_xref="taxon:1392"
                     /plasmid="virulence plasmid PX01"
     CDS             complement(1127..1630)
                     /note="Hypothetical protein, Methanobacterium
                     thermoautotrophicum, (AE000831)"
                     /codon_start=1
                     /transl_table=11
                     /product="pX01-01"
                     /protein_id="AAD32305.1"
                     /db_xref="GI:4894217"
                     /translation="MEVLIF[illegible]ILIAVLIFINSVVKKHVPKWKGKAGEKLVKRMLSKL
                     DPKSYYVLHNVTVYTEY[illegible]IDHIVIAEIGVFVVETKNYEGWIYGSEKAAPWTQGI
                     FRKKSSFQNPFHQNYKHIKA[illegible]FWLIEQQLPCISMAAFHPKCSLKPVNVHSKEKHVKYL
                     HNKFAKQ"
     RBS             complement(1639..1644)
     RBS             2751..2755
     CDS             2760..3338
                     /codon_start=1
                     /transl_table=11
                     /product="pX01-02"
                     /protein_id="AAD32306.1"
                     /db_xref="GI:4894218"
                     /translation="MLFHI[illegible]FSIVQLVPLIVSIGIIYILLELIDIEWKIWFSYEGIF
                     AIIFLITPFQVFQNKKSI[illegible]NQHEKWFSKNIFLVSIFLVLFASIPILLKIAFILQM
                     TVILFLLEYLCLYIDRLPIHNLYRVKEIICDKKQLKIERIIPNNEEHGTHTIKTRDEQ
                     YFLYIAVYLSGILSPISEEVFEEYERNNLMGR"
     CDS             complement(3389..4024)
                     /codon_start=1
                     /transl_table=11
                     /product="pX01-03"
                     /protein_id="AAD32307.1"
                     /db_xref="GI:4894219"
                     /translation="MKFEF[illegible]IGVLIILVLMILLLVFLQAAIASVKVIFNKLKSNIVS
                     SSFKDGLNNKIKIIEKIDSF[illegible]SIIQKIEGLYAPAFKVGVNIFSISFVLGVIVMVMAL
                     ITKNQQWMKWATGSMIFSFI[illegible]MMSMFLSVYFLFSSDVFSFMNIFIDFLILFKATALIF
                     TPFMLIAGIRLRRIHESTKQPEYYRWANRVITGSCLLCMLAVLAPWLFANM"
     CDS             complement(4124..4[illegible])
                     /codon_start=1
                     /transl_table=11
                     /product="pX01-04"
                     /protein_id="AAD32308.1"
                     /db_xref="GI:4894220"
                     /translation="MNIN[illegible]NVVEKLEIIVRYEKKIVIVLRVSKSSSNIHAYMIFF
                     EFSEVKISNTF[illegible]EYIEN[illegible]STSVKYVIDRKLSQQEIEKQNHSFIMTIEVRFPGVIDDE
                     TIHFNTTYVKPPTIANKVFKLVGGLGNRLDSLKELHITVLNNNINPIEIEFYNENETY
                     VAVEKEEIANAIMGAYIKANIYYAADLYYAAEGNEIKLKSIKPKLAGDV"
     RBS             complement(4034..4039)
     RBS             complement(4660..4665)
     CDS             complement(4989..5319)
                     /codon_start=1
                     /transl_table=11
                     /product="pX01-05"
                     /protein_id="AAD32309.1"
                     /db_xref="GI:4894221"
                     /translation="MEIIYF[illegible]MYRIIEYVSIFFETEFVFFVIIRVADVKIMIFVFNT
                     TKVLEIEIGLIKVINMSNH[illegible]WIHIEGEILFENILLYMKLEVDSFYSYILVEWSNFI
                     LGNI"
     RBS             complement(5318..5323)
     CDS             complement([illegible]..[illegible])
                     /codon_start=1
                     /transl_table=11
                     /product="pX[illegible]"
                     /protein_id="AAD3[illegible].1"
                     /db_xref="GI:[illegible]222"
                     /translation="[illegible]SFKVLANI
                     GNVITPVFESI[illegible]FINV[illegible]GIM[illegible]GVRVKQNV
                     LEVDGEIEDIIHA[illegible]NKIIDGQY
                     EFDLFG"
     RBS             complement([illegible])
     CDS             complement([illegible])
                     /note="reverse transcriptase [illegible] (AF 11549);
                     reverse transcriptase [illegible] (AF 11549);
                     306/543 pos[illegible]"
                     /codon_start=1
                     /transl_table=11
                     /product="pX[illegible]"
                     /protein_id="AAD32[illegible].1"
                     /db_xref="GI:[illegible]223"
                     /translation="[illegible]YRAEQFGN
                     KRKVRKLQRLM[illegible]VKLFHLLKGY
                     SVRNIKAFPVK[illegible]FLSMSYGF
                     RPKRSAHDAMAN[illegible]PYSNAINQ
                     WLNACCIDNK[illegible]DGAMLYPD
                     SIIIVPYALL[illegible]DGFDFLGF
                     NFRQYPIKEQ[illegible]NPVITGIANY
                     WSPVVAKVIYC[illegible]SQDKWLLTD
                     PDNHKNQLIRM[illegible]NSKRKLAK
                     KIKYKCRVCNN[illegible]HALLEWYGN
                     GKQLPKVQAY[illegible]KY"
     RBS             complement([illegible])
     CDS             complement([illegible]..11541)
                     /codon_start=1
                     /transl_table=11
                     /product="pX[illegible]"
                     /protein_id="AAD3[illegible].1"
                     /db_xref="GI:[illegible]224"
                     /translation="M[illegible]HAMAINHHREI
                     QFAVLGANDSI[illegible]QFEFINIQ
                     TINKKKAIIAF[illegible]KWAKPIYEEMLT
                     KNLL[illegible]TGEKINKI
                     QDLNEYLLEYS[illegible]AHVATGAAKA
                     LQVQKGIIIQC[illegible]KKWAKEEIRH
                     LIPLALVHLIK[illegible]QMPLPYKQ
                     IAL[illegible]KDFIGSDL
                     DSKTNKNSVQ[illegible]LQKQLELEN
                     HVKHYLAK[illegible]IALDSLAKV
                     LKKVVAGIGIL[illegible]NIERTYYHD
                     GEVISNIASR[illegible]NPVELVNVPT
                     ILVDMSEEQK[illegible]NMIHYPNYY
                     LLVAIKELVW[illegible]GSTVKERD
                     VQPRI[illegible]FLVKVGLDC
                     ATRSR"
     RBS             complement(115[illegible]..11554)
     CDS             complement([illegible]..[illegible]40)
                     /codon_start=1
                     /transl_table=11
                     /product="pX[illegible]"
                     /protein_id="AAD3[illegible].1"
                     /db_xref="GI:[illegible]225"
                     /translation="M[illegible]ASQLGAGTHLE
                     VLEQFSVFHGV[illegible]SAAFQSLPEWKRF[illegible]"
     RBS             complement([illegible]47..[illegible]52)
     CDS             complement([illegible]..[illegible]47)
                     /codon_start=1
                     /transl_table=11
                     /product="pX[illegible]"
                     /protein_id="AAD3[illegible].1"
                     /db_xref="GI:[illegible]226"
                     /translation="M[illegible]IFDPCGGEG
                     TILRSFADETK[illegible]SHNYFPLIF
                     LNPPYNTELR[illegible]DISYYLAKN
```

```
                     YEEIGLMRFGDENEEFDLFKLCVFIGRKRATTKNDTYFNDPFANFCENMSDLDFVRKH
                     VSTLAQMVNHKDWTIPELFHLCNKLIFISPVDYKDAYKGIGKSEGIASLKKPLNPGNGY
                     LLQGEKTAAEPAKMPIASGQLGLLLIIGVADGLLGEGDTLHAVPGSENVYYEHSFEQL
                     EKTTKEVVLQKRRAKFITATPIGVVKELV"
     RBS             complement(12956. .12961)
     CDS             complement(13209. .13562)
                     /codon_start=1
                     /transl_table=11
                     /product="pXO1-11"
                     /protein_id="AAD32315.1"
                     /db_xref="GI:4894227"
                     /translation="MFDNQKSAAPVMSSESPAPVNQMPGCFEAEIILPDGKFQYFKP
                     HALNEKEATQFIALYIEGIQFMKVGYVILWPFVGQQSFYSGTKIVAEPKFPKILPKLK
                     DYFFDYEEVVYEPKK"
     RBS             complement(13574. .13579)
     RBS             14226. .14231
     CDS             14237. .14929
                     /codon_start=1
                     /transl_table=11
                     /product="pXO1-12"
                     /protein_id="AAD32316.1"
                     /db_xref="GI:4894228"
                     /translation="MLKNILICSISFIVFSLIFFVIPIQLDIMIKIVFVIFNGLGISI
                     PTIVAMPMQLKQTIKLDFDIKFMKKEITKKFITNFVGLLKLQKYIQGDLKPLSEHYIN
                     RESEMHGLYREFRDWKDKYNAYIIPINMVAEKLIDDYASLLNLLEELDPHIQKIVLSY
                     VFTACEEYLNAIPSLGGSSEKDKEVVKLLSRNIQKKWIVLSAGFEKVFKEVQEETYTE
                     FTFLVQKQNKRG"
     CDS             complement(15040. .15802)
                     /note="erythrocyte invasion and possible binding protein;
                     Rhoptry protein (2401 aa), Plasmodium yoelii (U36927);
                     520/1194 positive aa (43%)"
                     /codon_start=1
                     /transl_table=11
                     /product="pXO1-13"
                     /protein_id="AAD32317.1"
                     /db_xref="GI:4894229"
                     /translation="MAKKFEIFDIYKYEDIIKAAKIGNLIIFVGAGISMLLGLPSWKG
                     YAQEILEDAIEKGVIDSFIYEWLKEEDPKKILTICKMLFDEKEIIPKSANEIFKFEKN
                     EEYKAIYEKLYSMNAIYITTNYDECLDVLTTPQSDNEKVSNSMDLSEENKLTEQSSET
                     KVFIDHTELLEAELKNGNVIHIHGSVKKEESMLVTVQDYLERYGTISKQSHPEVSVFL
                     DQVFNSNYVVLFMGYGLEEFEILEYMLSKTKNPSHTISHYMLFSENEENPQMTNLYKK

alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-174-937-2 x AF065404   ..

 Align seq 1/1  to: AF065404  from: 1  to: 181654

     262 ArgLysAsnLeuLysSerLeuIle 269
         ||||||||||||||||||||||||
   66712 AGAAAGAATTTGAAAAGTCTTATA 66735
```

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 02:33:42 ; Search time 159.56 Seconds
(without alignments)
2233.102 Million cell updates/sec

Title: US-09-174-937-1
Perfect score: 3076
Sequence: 1 ccacgcgtccgcccacgcgt..........ggcatggtggtgcatgcctg 3076

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/1/ina/NA_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/7D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	303.4	9.9	3035	2	US-08-726-725-2	Sequence 2, Appli
c	2	290.8	9.5	4421	3	US-08-257-963B-9	Sequence 9, Appli
c	3	290.8	9.5	4421	5	PCT-US95-07201-9	Sequence 9, Appli
c	4	289	9.4	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	5	285.6	9.3	2923	1	US-08-480-449-1	Sequence 1, Appli
c	6	285.6	9.3	2923	4	US-08-660-542-1	Sequence 1, Appli
c	7	285.6	9.3	2923	4	US-08-980-107-1	Sequence 1, Appli
	8	281.6	9.2	8392	1	US-08-080-255-6	Sequence 6, Appli
	9	281.6	9.2	8392	5	PCT-US93-05857-6	Sequence 6, Appli
	10	281.4	9.1	6769	2	US-08-480-784-20	Sequence 20, Appl
	11	281.4	9.1	6769	2	US-08-483-553-20	Sequence 20, Appl
	12	281.4	9.1	6769	2	US-08-487-002-20	Sequence 20, Appl
	13	281.4	9.1	6769	2	US-08-480-554B-20	Sequence 20, Appl
	14	281.4	9.1	6769	2	US-08-499-011B-20	Sequence 20, Appl
	15	281.4	9.1	6769	5	PCT-US95-10202-20	Sequence 20, Appl
	16	281.4	9.1	6769	5	PCT-US95-10203-20	Sequence 20, Appl
	17	281.4	9.1	6769	5	PCT-US95-10220-20	Sequence 20, Appl
	18	269.8	8.8	8342	5	PCT-US94-04496-63	Sequence 63, Appl
	19	267.6	8.7	5543	4	US-08-687-080-101	Sequence 101, App
c	20	267.4	8.7	1442	3	US-08-471-557C-120	Sequence 120, App
c	21	267.4	8.7	1442	4	US-08-340-426D-120	Sequence 120, App
c	22	267.4	8.7	1442	4	US-08-450-673C-120	Sequence 120, App
	23	263.8	8.6	4421	3	US-08-257-963B-9	Sequence 9, Appli
	24	263.8	8.6	4421	5	PCT-US95-07201-9	Sequence 9, Appli
c	25	263.4	8.6	7620	1	US-07-767-135-1	Sequence 1, Appli
c	26	263.4	8.6	7620	1	US-07-841-652-1	Sequence 1, Appli
	27	262	8.5	11613	1	US-08-484-044-10	Sequence 10, Appl
	28	254.8	8.3	246240	3	US-08-724-394A-20	Sequence 20, Appl
	29	254.8	8.3	246240	3	US-08-724-394A-21	Sequence 21, Appl
	30	254.8	8.3	246240	3	US-08-724-394A-22	Sequence 22, Appl
	31	250.4	8.1	2589	6	5212286-1	Patent No. 5212286
	32	245.8	8.0	3865	2	US-08-832-993-48	Sequence 48, Appl
	33	245.8	8.0	3865	3	US-08-832-877-48	Sequence 48, Appl
c	34	244.4	7.9	8174	1	US-07-914-281-5	Sequence 5, Appli
c	35	244.4	7.9	8174	1	US-08-091-240-5	Sequence 5, Appli
c	36	244.4	7.9	3373	1	US-08-273-411-2	Sequence 2, Appli
c	37	244.4	7.9	8174	2	US-08-525-059A-5	Sequence 5, Appli
c	38	244.4	7.9	8174	4	US-08-990-721-5	Sequence 5, Appli
c	39	244.4	7.9	8174	5	PCT-US91-10899-5	Sequence 5, Appli
c	40	242	7.9	3267	3	US-08-257-963B-12	Sequence 12, Appl
c	41	242	7.9	3267	5	PCT-US95-07201-12	Sequence 12, Appl
c	42	242	7.9	22481	5	PCT-US95-17211-43	Sequence 43, Appl
c	43	241.8	7.9	246240	3	US-08-724-394A-20	Sequence 20, Appl
c	44	241.8	7.9	246240	3	US-08-724-394A-21	Sequence 21, Appl
c	45	241.8	7.9	246240	3	US-08-724-394A-22	Sequence 22, Appl

ALIGNMENTS

```
RESULT 1
US-08-726-725-2/c
; Sequence 2, Application US/08726725
; Patent No. 5773290
;  GENERAL INFORMATION:
;    APPLICANT:  Gould, Michael N.
;    APPLICANT:  Chen, Kai-Shun
;    TITLE OF INVENTION:  MAMMARY GLAND-SPECIFIC PROMOTERS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,725
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  960296.93863
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5709
;      TELEFAX:  (414) 271-3552
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3035 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-726-725-2

  Query Match             9.9%;  Score 303.4;  DB 2;  Length 3035;
  Best Local Similarity   82.5%;  Pred. No. 1e-65;
  Matches  397;  Conservative    0;  Mismatches   71;  Indels   11;  Gaps    3;

Qy       2615 gctgggccaggcacgat[illegible] 2674

Db       1465 [illegible] 1406

Qy       2675 ggatcacctgaggtgaggagttt[illegible] 2734

Db       1405 GGATCAC--GAGGTCAGGAGTT[illegible] 1348

Qy       2735 actaaaaatacaaaaattag[illegible] 2794
```

```
Db    1347 ACTAAAAATACAAAAATTAGCAGGGCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCGG 1288

Qy    2795 gaggctgaggcatgagaatcacttaaacctgaagaggcagaggttacagtgagccaagatc 2854
           |||||||||| |||| ||| ||| || |||| ||| || |||||||||| |||||
Db    1287 GAGGCTGAGGCAGAAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATC 1228

Qy    2855 gtgccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaataaaaaa 2914
             ||| |||| | |||||||   ||| |||| | | ||||||| |||||||| ||||||
Db    1227 AGGCCTTTGCACTTTATGCTGG--AACATAGTGATATTCTGTCTTAAAAAAAAAAAAAAA 1170

Qy    2915 -------taatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgag 2967
                  | |||| |||||||||  ||| |||  |||||||| |||||| ||||
Db    1169 GGTATCCTCGGGTTGGGTGCAGTGGCTCTTGCCTGTAATTCCAGCACTTTGGGAGACTGAG 1110

Qy    2968 gtgggtggatcacctgaggtcaggagttcgagaccagcctggtgaacatggcgaaaccccg 3027
           | |||||||| |||||||||||| |||||||||||||||   ||||||| |||||||
Db    1109 GCGGGTGGATAACCTGAGGTCAGGTGTTCAAGACCAGCCTGACCAACATGGTGAAACCCT 1050

Qy    3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||| ||||||||||||| |||||   ||| |||||| |||||||||
Db    1049 GTCTCTACTAAAAATACAAAGATTAGTTGGGCATGGTGGCGCATGCCTG 1001


RESULT   2
US-08-257-963B-9/c
; Sequence 9, Application US/08257963B
; Patent No. 5840686
;  GENERAL INFORMATION:
;    APPLICANT:  Chader, Gerald J.; Becerra, S.
;    APPLICANT:  Patricia; Schwartz, Joan P.;
;    APPLICANT:  Taniwaki, Takayuki
;    TITLE OF INVENTION:  PIGMENT EPITHELIUM
;    TITLE OF INVENTION:  DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
;    TITLE OF INVENTION:  BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
;    TITLE OF INVENTION:  AND EXPRESSING THE PROTEIN
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  08/257,963B
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/952,796
;      FILING DATE:  24-SEPT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOROTHY R. AUTH
;      REGISTRATION NUMBER:  36434
;      REFERENCE/DOCKET NUMBER:  22154.26US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4421 Base Pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Human
;    FEATURE:
;      NAME/KEY:  JJ101
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  7.1 kb Bam HI
;      OTHER INFORMATION:  fragment Derived from human placental
;      OTHER INFORMATION:  genomic DNA
US-08-257-963B-9


  Query Match             9.5%;  Score 290.8;  DB 3;  Length 4421;
  Best Local Similarity   84.0%;  Pred. No. 1.5e-62;
  Matches  388;  Conservative    0;  Mismatches   67;  Indels    7;  Gaps    5;

Qy   2619 ggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcgggcgg 2677
          ||||||| | | |||||||||||||| ||||||| |||||||||| || | |||
Db   4089 GGCCAGGTATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTAAGGTGGGTGGA 4030

Qy   2678 tcacctgaggtgaggagtttgagaccagcctgtccaacatggtgaaaccctgtctctact 2737
          ||||||| || |||||||  ||| |||||||||||||||||||||||| |||| | |
Db   4029 TCACCTGATGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTTTACT 3970

Qy   2738 aaaaaatac-aaaaattagccaggcgtggtggcaggtgcctgtaatcccagctactcggga 2796
          |||||||| ||||| |||||||||| |||| || |||| |||||||||| | | |
Db   3969 AAAAATACAAAAAACTAGCCAGGGCGGGTGGTGGGCACCTGTAATCCCAGCTACTCGGGA 3910

Qy   2797 ggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagccaagatcgt 2856
          |||||||||| ||||||| ||| || | |  |||| |||||| |||||||||| ||||||
Db   3909 GGCTGAGGCAGGAGAATCTCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAGATCGC 3850

Qy   2857 gccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaataaaaaata 2916
          |||||||| |||||||  || ||| | | |||||  |||||||||| | | | | |
Db   3849 TCCACTGCACTCCAGCCCAGGTGACAGTGTGAGACTCCATCTCAAAAAAAA-AAATAAATA 3791

Qy   2917 atcagggcacagtggctcatgcctgtaatcccagcactttgggaggccgaggcgggcgga 2976
          | |||  | |||||||| ||||| |||||||||||| |||||||| |||| ||   |
Db   3790 GGCCGGGTGCGGTGGCTCAAGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGA 3731

Qy   2977 tcacctgaggtcaggagttcaagaccagcctgatgaacatggcgaaaccccgtctctaat 3036
          ||||  |||||||||||| || |||| | |||||  |||| || |||||||| ||||
Db   3730 TCAC--GAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACT 3673

Qy   3037 aaaaatac--aaaaattagccgggcatggtggtgcatgcctg 3076
          ||||||||  |||||||| ||||| |||||| | | |||||
Db   3672 AAAAATACAAAAAAAATTAGTCGGGCGTGGTGGCGGACGCCTG 3631


RESULT   3
PCT-US95-07201-9/c
; Sequence 9, Application PC/TUS95/07201
;  GENERAL INFORMATION:
;    APPLICANT:  Chader, Gerald J.; Becerra, Sofia
;    APPLICANT:  Patricia; Schwartz, Joan P.;
;    APPLICANT:  Taniwaki, Takayuki
;    TITLE OF INVENTION:  PIGMENT EPITHELIUM
;    TITLE OF INVENTION:  DERIVED FACTOR: CHARACTERIZATION GENOMIC
;    TITLE OF INVENTION:  ORGANIZATION AND SEQUENCE OF THE PEDF GENE
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07201
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  08/367,841
;      FILING DATE:  30-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/257,963
;      FILING DATE:  07-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/952,796
;      FILING DATE:  24-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOROTHY R. AUTH
;      REGISTRATION NUMBER:  36434
;      REFERENCE/DOCKET NUMBER:  20264126PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4421 Base Pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Human
;    FEATURE:
;      NAME/KEY:  JT1
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  7.1 kb Bam HI
;      OTHER INFORMATION:  fragment Derived from human placental
;      OTHER INFORMATION:  genomic DNA; Also referred to as JT101
PCT-US95-07201-9


  Query Match             9.5%;  Score 290.8;  DB 5;  Length 4421;
  Best Local Similarity   84.0%;  Pred. No. 1.5e-62;
  Matches  388;  Conservative    0;  Mismatches   67;  Indels    7;  Gaps    5;

Qy    2619 ggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcgga 2677
           ||||||| | | |||||||||||||||||||||:| ||||||||||||  |||  ||||||
Db    4089 GGCCAGGTATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTAAGGTGGGCGGA 4030

Qy    2678 tcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctact 2737
           |||||||| || |||||||  ||| |||||||||||||||||||||||||||| |||| ||||
Db    4029 TCACCTGATGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTTTACT 3970

Qy    2738 aaaaatac-aaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggga 2796
           |||||||| ||||| ||||||||||||||||  ||  ||||||||||||||||||||||||
Db    3969 AAAAATACAAAAAACTAGCCAGGCGTGGTGGCTGGGCACCTGTAATCCCAGCTACTTGGGA 3910

Qy    2797 ggctgaggcatgagaatcacttaaacctgaaaggcagaggttacagtgagccaagatcgt 2856
           ||||||||| |||||||  ||| ||||   |:|:|| |||||| |||||||| ||||||
Db    3909 GGTTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGC 3850

Qy    2857 gccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaataaaaaata 2916
           ||||||| |||||||  || |||| |   |||||  |||||||||||| ||| || | |
Db    3849 TCCACTGCACTCCAGCCCAGGTGACAGTGTGAGACTCCATCTCAAAAAAAA-AAATAAGA 3791

Qy    2917 atcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtgga 2976
           | ||| | ||||||||  ||||||||||||||||||  |||||||| |||| || ||
Db    3790 GGCTGGGTGTGGTGGCTCAAGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCAGA 3731

Qy    2977 tcacctgaggtcaggagttcaagaccagcctggccaacatggcgaaaccccgtctctaat 3036
           |||| ||||||||||| || |||||| ||||  |||| || ||||||||||||||| |
Db    3730 TCAC--GAGGTCAGGAGATCCAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACT 3673

Qy    3037 aaaaatac--aaaaattagccgggcatggtggtgcatgcctg 3076
           ||||||||  |||||||| ||||| || || | | |||||
Db    3672 AAAAATACAAAAAAATTAGTCGGGCGTGGTGGTGGGACGCCT3 3631


RESULT   4
US-08-323-443B-1/c
```

```
; Sequence 1, Application US/08323443B
; Patent No. 5654170
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W.
;    APPLICANT:  LANDES, GREGORY M.
;    APPLICANT:  BURN, TIMOTHY C.
;    APPLICANT:  CONNORS, TIMOTHY D.
;    APPLICANT:  DACKOWSKI, WILLIAM R.
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESSEE:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,443B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  [illegible]/DA462
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  PKD1 GENOMIC
US-08-323-443B-1


  Query Match             5.4%;  Score 164;  DB 1;  Length 31571;
  Best Local Similarity   82.4%;  Pred. No. 9.3e-62;
  Matches  380;  Conservative    0;  Mismatches   72;  Indels    6;  Gaps    4;

Qy    2611 aaaggctgggccaggcacgatgctcatgcctgtaatcccagcactttgggaggccgagg 2669
           |||| || |||  |||| || | | | | | | | | | | | | | | | | | | | |
Db     639 AAAGCCTAGGCCAGGCCACAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 580

Qy    2670 caggcggatcacctgaggtgaggagttcgagaccagcctggccaacatggtgaaaccccg 2729
           |||||| ||||||||| | | | | | | | | | | | | | | | | | | | | | | |
Db     579 CAGGCGGATCACATGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCG 520

Qy    2730 tctctactaaaaatacaaaaaattagccgggcgtggtggtgcgcgcctgtaatcccagcta 2789
           |||||||||||||||||||| | | | | | | | | | | | | | | | | | | | | |
Db     519 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGGGCGCCTGTAATCCCAGCTA 460

Qy    2790 cttgggaggctgaggcaggagaatcgcttgaacctgggaggcggaggttgcagtgagccg 2849
           || |||||||||||| |||| | | | | | | | | | | | | | | | | | | | | |
Db     459 CCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCG 400

Qy    2850 agatcgtgccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaata 2909
           |||  | |||||||| |||| | | | | | | | | | | | | | | | | | | | |
Db     399 AGGTTACGCCGACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAGA-A 341
```

```
Qy   2910 aaaaataatcaaggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggt 2969
          ||||| | | |||||||||||||||| || ||||||||||||| |||||||| ||||
Db    400 AAAAAAAGGCCAGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCGGAGGC 341

Qy   2970 gggtggatcacctgaggtcaggagttcgagaccagcctggtgaacatggcgaaaccccgt 3029
          |||||||||| |||||| ||| |   | | | ||||| | |||| ||||||||||
Db    340 AGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGT 284

Qy   3030 ctctaataaaaatacaaaaattagccgggcatggtggtgca 3070
          |||| ||||||||||||||||| || ||||||||| ||
Db     283 CTCTACTAAAAATACAAAAATTGGCTGGGCATGGTGGCACA 243


RESULT   5
US-08-480-449-1/c
; Sequence 1, Application US/08480449
; Patent No. 5888947
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    TITLE OF INVENTION:  MACROPHAGE DERIVED CHEMOKINE
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,449
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  27866/32779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..298
US-08-480-449-1


  Query Match             9.3%;  Score 285.6;  DB 1;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No. 2.5e-61;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy   2573 agaagcactgagctcccctgtctctggaaatatttaagaaaaggctgggccaggcacgatg 2632
          | | | | | | | | ||| | || || ||| | ||| || |||
Db   1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy   2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
          |||||||||||||||| ||| |||||||||| | ||| ||| ||||||| |||||| ||
Db   1805 GCTCACGGCTGTAATCTCAGCACTTTGGGAGGCACAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy   2692 gagtttgagaccagcctggccaacatggtgaaacctctctctactaaaaatacaaaaat 2751
          ||||| ||| |||||||||||| | |||||||||| ||||||||||||||||||||
Db   1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-CTCTACTAAAAATACAAAAAT 1687

Qy   2752 tagccaggcgtggtggcaggtgcctgtaatcccagctactcgggaggctgaggcaggaga 2811
          ||||||| ||| ||||||    ||| |||||||||||||||||||||| | | | |
Db   1686 TAGCCAGCCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGA 1627

Qy   2812 atcacttaaacctgagaggcagaggttgcagtgagccgagatcgtgccactgcactccag 2871
          ||||||| ||||   |||| |||| | |||||||| |||||  ||||||| ||||||||
Db   1626 ATCACTTGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy   2872 cctgggcgacagagcaagactctgtctcaaaaaaaaataaaaaat--- -          --- 2916
          ||||||||||||||  |||| |||    ||||||||| |||||
Db   1566 CCTGGGCGACAGAGTGAGACCCTGCCATAAAAAAAAAAAAAAAAGTTGGGGGGGGCAGTGG 1507

Qy   2916 ----------aatcagggcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                     | | ||||| |||||||| ||||||||||||||||||| | | |
Db   1506 GGAGAGGGAAGAGGCCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447

Qy   2965 gaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtggaaacatgcgaaac 3024
          |||||||| |||| |||||||| |||||||| |||||||||||  |||||||| | |
Db   1446 GAGGTGGGCGGATTACCTGAGGACAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAAC 1387

Qy   3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          | |||||||| ||||||||||||||||||| |||| |||| |    || |
Db   1386 CTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTAGCAGGAACCTG 1335


RESULT   6
US-08-660-542-1/c
; Sequence 1, Application US/08660542
; Patent No. 5932703
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    TITLE OF INVENTION:  MACROPHAGE DERIVED CHEMOKINE AND CHEMOKINE
;    TITLE OF INVENTION:  ANALOGS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/660,542
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/558,658
;      FILING DATE:  16-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,620
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  27866/33318
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..298
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  92..298
US-08-660-542-1


  Query Match             9.3%;  Score 285.6;  DB 1;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No:  2.5e-61;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy    2573 agaagcactgagctcgctgtctctaaaaaaaatttaagaaaaaa tgagc aaaaaaaaa 2632
           | | | |  || | |  |  |||   | ||  || ||| ||  |||  ||  |||
Db    1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy    2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
           |||||||||||||| ||| |||||||||| | ||| ||| ||||||| |||||| ||
Db    1805 GCTCACGCCTGTAATCTCAGCACTTTGGGAGGCAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy    2692 gagtttgagaaacagcctggccaacatggtgaaacctcatctctactaaaaatacaaaaat 2751
           ||||   ||| ||||||||||| |||||||  |||  |||||||||||||||||||||
Db    1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-GTCTCTACTAAAAATACAAAAAT 1687

Qy    2752 tagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgaga 2811
           |||||||| ||| |||||   ||||||| | ||||||||||||||||| ||   ||||
Db    1686 TAGCCAGGCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGGGGTGGGAGA 1627

Qy    2812 atcacttaaacctgagaggcagaagttgcagtgagccaagatcgtgccactgcattccag 2871
           |||||||| |||  ||||  |||||   |||||| |||||  ||||||||| |||||
Db    1626 ATCACTTGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy    2872 cctgggcgacagagcgagactctgtctcaaaaaaaataaaaaat·················· 2916
           ||| ||| |||| | | | || ||||||
Db    1566 CCTGGGCGACAGAGTGAGACCCTGTCATAAAAAAAAAAAAAAAGTTGGGGGGGGGCAGTGG 1507

Qy    2916 ----------aatcaggcgcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                     |||| |||||| ||||||||||||||| |||||||||
Db    1506 GGAGAGGGAAGAGGCCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447

Qy    2965 gaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaac 3024
           |||||||  ||| |||| ||||| | ||||||||||| |||||||  || |
Db    1446 GAGGTGGGCGGATTACCTGAGGACAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAGC 1387

Qy    3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||| | ||||||||| |||||||  | ||||       ||||
Db    1386 CCCGTCTCTACTAAAAATACAAAAATTAGTTGGGCGTGGTGGTAGCAGGAACCTG 1335


RESULT  7
US-08-989-107-1/c
; Sequence 1, Application US/08989107
; Patent No. 5992828
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Raport, Carol J.
;    TITLE OF INVENTION:  MACROPHAGE DERIVED CHEMOKINE (MDC) AND
;    TITLE OF INVENTION:  CHEMOKINE ANALOGS AND ASSSAY TO IDENTIFY MODULATORS OF MDC ACT
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/989,107
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/559,658
;      FILING DATE:  16-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/479,611
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,151
;      REFERENCE/DOCKET NUMBER:  17810/38018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20..298
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  92..298
US-08-989-107-1


  Query Match             9.3%;  Score 285.6;  DB 4;  Length 2923;
  Best Local Similarity   75.9%;  Pred. No:  2.5e-61;
  Matches  404;  Conservative    0;  Mismatches   99;  Indels   29;  Gaps    3;

Qy    2573 agaagcactgagctcgctgtctctaaaaaaaatttaagaaaaaa tgagc aaaaaaaaa 2632
           | | | |  || | |  |  |||   | ||  || ||| ||  |||  ||  |||
Db    1865 AAACGAAGAAAGGTACAGGAGTCTCCTATTAGGAAAAAAAGGGGGAGGCTGGGTGTGATA 1806

Qy    2633 gctcacgcctgtaatcccag-actttgggaggccgaggcaggcggatcacctgaggtgag 2691
           |||||||||||||| ||| |||||||||| | ||| ||| ||||||| |||||| ||
Db    1805 GCTCACGCCTGTAATCTCAGCACTTTGGGAGGCAAGGAAGGTGGATCACTTGAGGTCAG 1746

Qy    2692 gagtttgagaaacagcctggccaacatggtgaaacctcatctctactaaaaatacaaaaat 2751
           ||||   ||| ||||||||||| |||||||  |||  |||||||||||||||||||||
Db    1745 GAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCT-GTCTCTACTAAAAATACAAAAAT 1687

Qy    2752 tagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggctgaggcatgaga 2811
           |||||||| ||| |||||   ||||||| | ||||||||||||||||| ||   ||||
Db    1686 TAGCCAGGCGTAGTGGCACACACCTGTAATCCCAGCTACTTGGGAGGCTGGGGTGGGAGA 1627

Qy    2812 atcacttaaacctgagaggcagaagttgcagtgagccaagatcgtgccactgcattccag 2871
           |||||||| |||  ||||  |||||   |||||| |||||  ||||||||| |||||
Db    1626 ATCACTTGAACCCAGGAGGTCGAGGTTGTAGTGAGCCGAGATCACGCCACTGCACTCCAG 1567

Qy    2872 cctgggcgacagagcgagactctgtctcaaaaaaaataaaaaat·················· 2916
           ||| ||| |||| | | | || ||||||
Db    1566 CCTGGGCGACAGAGTGAGACCCTGTCATAAAAAAAAAAAAAAAGTTGGGGGGGGGCAGTGG 1507

Qy    2916 ----------aatcaggcgcacagtggctcatgcctgtaatcccagcactctgggaggct 2964
                     |||| |||||| ||||||||||||||| |||||||||
Db    1506 GGAGAGGGAAGAGGCCAGGCACTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCT 1447

Qy    2965 gaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaac 3024
           |||||||  ||| |||| ||||| | ||||||||||| |||||||  || |
Db    1446 GAGGTGGGCGGATTACCTGAGGACAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAGC 1387

Qy    3025 cccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||| | ||||||||| |||||||  | ||||       ||||
Db    1386 CCCGTCTCTACTAAAAATACAAAAATTAGTTGGGCGTGGTGGTAGCAGGAACCTG 1335
```

```
Db   1386 CTGGTCTCTACTAAAAAATACAAAAAATTAGCTGGGCGTGGTAGCAGGAAGCTG 1335


RESULT   8
US-08-080-255-6
; Sequence 6, Application US/08080255
; Patent No. 5487970
;  GENERAL INFORMATION:
;    APPLICANT:  Powley, Janet D.
;    APPLICANT:  Diaz, Manuel O.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/080,255
;      FILING DATE:  19930617
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  APCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8392 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-080-255-6


  Query Match             9.2%;  Score 281.6;  DB 1;  Length 8392;
  Best Local Similarity   78.8%;  Pred. No. 3.6e-60;
  Matches  361;  Conservative    0;  Mismatches   94;  Indels    3;  Gaps    2;

Qy    2620 gccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcggat 2678
           | || |     | ||||||||||||| ||||| | ||||||||| ||||| |||||| |||
Db    1179 GGCAAGGCGCACGGCTCACGCCTGTGATCCCAGCACTTTGGGAAGCCGAAGCAGGCAGAT 1238

Qy    2679 cacctgaggtgaggagtttgagaaacagcctggccaacatggtgaaaacctcatctactaa 2738
           ||| |||||| ||||||| |||| ||||||| |||||||||||||||||   ||||||||
Db    1239 CACTTGAGGTCAGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTA 1298

Qy    2739 aaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggagg 2798
           |||||||||||||||| | ||||| | ||||| |||||||| ||||||||||||||||
Db    1299 AAAATACAAAAATTAGCCGGTTGTGGTAGTGGGTGCTTGTAATCCTAGCTACTTGGGAGG 1358

Qy    2799 ctgaggcatgaaaatcacttaaacctgaaggcagaggttacagtgagccaagatcgtgc 2858
           |||||||| | |||| | | |||||| |||  |||| | |||||||| |||| |
Db    1359 CTGAGGCAGGGGAATTGCCTGAACCTGGGAGGTGGAGGCTGCAGTGAGCCGAGATTGCAT 1418

Qy    2859 cactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataaaaaataat 2918
           | ||| || ||||||| | ||||||| || ||  || |||||||||| ||||||
Db    1419 CATTGCACTCTAGCCTGGACAACAGAGCTAGACTCCATCCCAAAAAAAAAAAAAAAGTA 1478

Qy    2919 caggccaggtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtggatc 2978
```

```
Db   1479 GCCGGGCAGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATC 1538

Qy    2979 acctgaggtcaggagttcaagaccagcctggtgaaacatggcgaaaccccgtctctaataa 3038
           |  ||||| |||||| ||.|||||| ||||| |||| || |||||||| ||||| |.
Db    1539 A--TGAGGCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAA 1596

Qy    3039 aaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||||||||| |  | |  | | |  |||||
Db    1597 AAATACAAAAAATTAGCCCGGCGAGGTGCGGGCGCCTG 1634


RESULT   9
PCT-US93-05857-6
; Sequence 6, Application PC/TUS9305857
;  GENERAL INFORMATION:
;    APPLICANT:  Board of Regents
;    APPLICANT:  The University of Texas System
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  GENE REARRANGEMENTS AND TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05857
;      FILING DATE:  19930617
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/900,688
;      FILING DATE:  17/06/92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  APCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-05857-6


  Query Match             9.2%;  Score 281.6;  DB 5;  Length 8392;
  Best Local Similarity   78.8%;  Pred. No. 3.6e-60;
  Matches  361;  Conservative    0;  Mismatches   94;  Indels    3;  Gaps    2;

Qy    2620 gccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcggat 2678
           | || |     | ||||||||||||| ||||||| |||||||| |||||| |||
Db    1179 GGCAAGGCGCACGGCTCACGCCTGTGATCCCAGCACTTTGGGAAGCCGAAGCAGGCAGAT 1238

Qy    2679 cacctgaggtgaggagtttgagaaacagcctggccaacatggtgaaaacctcatctactaa 2738
           ||| |||||| ||||||| |||| ||||||||||||||||||||||||||||||||||
Db    1239 CACTTGAGGTCAGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTA 1298

Qy    2739 aaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggc 2798
           |||||||||||||||| | ||||| | ||||| |||||||| |||||||||||||||||
Db    1299 AAAATACAAAAATTAGCCGGTTGTGGTAGTGGGTGCTTGTAATCCTAGCTACTTGGGAGG 1358
```

```
Qy   2799 ctgaggcatgagaatcacttaaacctgaggaggcagaggttacagtgagccaagatcgtgc 2858
          ||||||||| | |||| | | |||||| | |||| |||| | |||||||||| |||| |
Db   1359 CTGAGGCAGGGGAATTGCCTGAACCTGCGACCCGGAGGCTGCAGTGAGCCGAGATTGCAT 1418

Qy   2859 cactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataaaaaataat 2918
          || ||| || ||||||| | ||||||| |||.|| || |||||||||| ||||||
Db   1419 CATTGCACTCIAGCCTGGACAACAGAGCTAGACTCCATCCCAAAAAAAAAAAAAAAAGTA 1478

Qy   2919 cagggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtggatc 2978
          ||    |||||||| |||||||||||||| ||| |||||||| |||| ||| ||||
Db   1479 GCCGGGCAGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATC 1538

Qy   2979 acctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccgtctctaataa 3038
          |  ||||| |||||| || |||||| |||||  |||| || ||||||| ||||||| |||
Db   1539 A--TGAGGGCAGGAGAICGAGACCATCCTGGCTAACACGGTGAAACCCTGTCTCTACTAA 1596

Qy   3039 aaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          |||||||||| |  |  |  |  | |  ||||
Db   1597 AAATACAAAAAATTAGCCGGGCGAGGTGGGCGCGCCTG 1634


RESULT 10
US-08-480-784-20
; Sequence 20, Application US/08480784
; Patent No. 5693473
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,784
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109947
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-480-784-20


  Query Match             9.1%;  Score 281.4;  DB 2;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-61;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy   2619 gggccaggcacgatggctcacgcctgtaatcccag actttgggaggccgaggcgggcgg 2676
          ||||  |||  | |||| | | | | | | | | | | | | | | | | | | | |
Db     32 GGGCTGGGCGTAGTAGCTTGCAGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAG 91

Qy   2677 atcacctgaggtgaggagtttgagaccagcctggc agcatggtgaaaccctgtctctac 2736
          || |  |||||| |||||||| | | | | | | | | | | | | | | | | | | |
Db     92 ATAATTTGAGGTCAGGAGTTTGAGACCAGCCTGGCTAACATGGTGAAACCCTGTCTCTAC 150

Qy   2737 taaaaatacaaaaattagccgggtgtggtggcacgtacctgtaatcccagctactcgggga 2796
          ||||||||||| |||||||| | | | | | | | | | | | | | | | | | | |
Db    151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy   2797 ggctgaggcatgagaatcacttgaacctgggaggcagaggttg gttacagtgagccgaga 2856
          ||| |||||  |||| | | | | | | | | | | | | | | | | | | | | |
Db    211 GGCGGAGGCAGAAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCGGTGAGCCGAGATCGT 270

Qy   2857 gccactgcattccagcctgggcgac agag agactctgtctcaaaaaaaaaaaaaaaa-- 2915
          ||| |||||| |||||||||||  | | | | | | | | | | | | | | | | |
Db    271 GCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAGT 330

Qy   2915 -------taatcagggcacagtggctcatgcctgtaatcccagcactttgggaggctgag 2957
                 | | |||| ||| | | | | | | | | | | | | | | | | | | |
Db    331 AATTGTTCTAGCTGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAG 390

Qy   2969 gtgggtggatcacctgaggtcaggagttcaagaccagcctggccaacatggtgaaaccccg 3027
          | |||||||| | ||| | | | | | | | | | | | | | | | | | | | |
Db    391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCC 448

Qy   3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          || |||  |||||||||||||| | | | | | | | | | | | | |
Db    449 ATCGCTACAAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGTGCCTG 497


RESULT 11
US-08-483-553-20
; Sequence 20, Application US/08483553
; Patent No. 5709999
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,553
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/288,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  21884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-483-553-20

  Query Match             9.1%;  Score 241.4;  DB 2;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-60;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy   2619 gggccaggcaggatggctcacgcctgtaatcccag-actttaggaggccgaggcaggcgg 2676
          |||| ||| |  | ||| | ||||||||||||| |||| ||||||| |||||||| |
Db     32 GGGCTGGGCTAGTAGCTTGCACCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGCAG 91

Qy   2677 atcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctac 2736
          || |  |||||| |||||||||||| |||||||||||||||||||||||||||| |||||||||
Db     92 ATAATTTGAGCTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACC-CATCTCTAC 150

Qy   2737 taaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggga 2796
          ||||||||||||||||||  || |||||||| || ||||||||||||||||||| ||||
Db    151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy   2797 ggctgaggcatgagaatcacttaaacctgaggaggcagaggttacagtgagccaagatcgt 2856
          ||| || |||  |||||  ||| ||||||  || || |||||| ||| ||||||||||||
Db    211 GGCGGAGGCAGAAGAATTGCTTGAACCTAGGAGCGGGAGGTTGCAGCGAGCCAAGATCGC 270

Qy   2857 gccattgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaataaaaaa-- 2915
          |||||||||| |||||||||||| |||| |||||||||||||||| | | | | | |
Db    271 GCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAGAAAAAAAAGT 330

Qy   2915 -------taatcagggcacagtggctcatgcctgtaatcccagcactctgggaggccaaag 2967
                 | | |||| ||||||||| ||||||||||||||||||| |||||||| ||
Db    331 AATTGTTCTAGCTGGGCGCAGTGGCTCTTGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 390

Qy   2968 gtgggtggatcacctgaggtca-ag-ttcaagaccagcctggtcaacatagtgaaaccccg 3027
          | |||||||| |  ||| | |||||||||||||||||| | || ||| |||||||||
Db    391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGTGGTGAAACCCC 448

Qy   3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          || ||| |||||||||||||||||| |||||||||| | || ||
Db    449 ATCGCTACAAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCATG 497


RESULT  12
US-08-487-002-20
; Sequence 20, Application US/08487002
; Patent No. 5710001
;  GENERAL INFORMATION:
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Emi, Mitsuru
;    APPLICANT:  Nakamura, Yusuke
;    APPLICANT:  Durocher, Francine
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,002
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/288,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  21884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-487-002-20


  Query Match             9.1%;  Score 281.4;  DB 2;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-60;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy    2618 gggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcgg 2676
           ||||  ||||    | |||  | ||||||||||||| |||| ||||||| |||||||| |
Db      82 GGGCTGGGGTCATAGTAGCTTGCACCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGCAG 91

Qy    2677 atcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctac 2736
           || | |||||| |||||||||| || | |||||||||||||| ||||||||| |||||||||
Db      92 ATAATTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCC-CATCTCTAC 150

Qy    2737 taaaaatacaaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttggga 2796
           | |||| ||||||||| || |||| | || |||||||||||||||| ||||
Db     151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy    2797 ggctgaggcatgagaatcacttaaacctgggaggcagaggttacagtgagccaagatcgt 2856
           || ||| | ||||  ||| ||||| |  | |||||  ||| ||||||||||||
Db     211 GGTGGAGGTTACAAGAATTGCTTGAACCTAAGAGGCGGAGGTTGCAGTGAGCCAAGATCGC 270

Qy    2857 gccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaataaaaaa-- 2915
           ||||||||| ||||||||||||    ||||  ||||||||||||||| | ||||||
Db     271 GCCACTGCACTCCAGCCTGGGCGGTAGAGCGAGACTCTGTCTCAAAAAAGAAAAAAAAGT 330

Qy    2915 ------taatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgag 2967
               |  ||||||||||| | | ||||||||||||| |||||||  ||
Db     331 AATTGTTCTAGCTGGGCGCAGTGGCTCTTGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 390

Qy    2968 gtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccg 3027
           | |||||||| | |  |||  | |||||||  || |||  |||| ||||||||
Db     391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGTGGTGAAACCCC 448

Qy    3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | | || |||||||||||||||  |  |||||| |   || ||
Db     449 ATCGCTACAAAAAATACAAAAATTAGCCACGCATGGTGGCGTGCGGCATG 497


RESULT  13
US-08-483-554B-20
; Sequence 20, Application US/08483554B
; Patent No. 5747282
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,554B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,815
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24944-104-47
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-483-554B-20


  Query Match             9.1%;  Score 281.4;  DB 2;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-60;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy    2618 gggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcgg 2676
Db      82 GGGCTGGGGTCATAGTAGCTTGCACCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGCAG 91

Qy    2677 atcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctac 2736
Db      92 ATAATTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCC-CATCTCTAC 150

Qy    2737 taaaaatacaaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttggga 2796
Db     151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy    2797 ggctgaggcatgagaatcacttaaacctgggaggcagaggttacagtgagccaagatcgt 2856
Db     211 GGTGGAGGTTACAAGAATTGCTTGAACCTAAGAGGCGGAGGTTGCAGTGAGCCAAGATCGC 270

Qy    2857 gccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaataaaaaa-- 2915
Db     271 GCCACTGCACTCCAGCCTGGGCGGTAGAGCGAGACTCTGTCTCAAAAAAGAAAAAAAAGT 330

Qy    2915 ------taatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgag 2967
Db     331 AATTGTTCTAGCTGGGCGCAGTGGCTCTTGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 390

Qy    2968 gtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccg 3027
Db     391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGTGGTGAAACCCC 448
```

```
Qy   3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          || ||| ||||||||||||||||||| |||||||| |   || ||
Db    449 ATCGCTACAAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCATG 497


RESULT  14
US-08-488-011B-20
; Sequence 20, Application US/08488011B
; Patent No. 5753441
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,011B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109847-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      .
;      .
;      .
US-08-488-011B-20


  Query Match             9.1%;  Score 281.4;  DB 0;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-60;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy   2618 gggccaggcacgatggctcacg[illegible] 2676
          ||||  |||   | |||  | ||||||||||||| |||| |||||| |||||||| |
Db     32 GGGCTGGGCGTAGTAGCTTGCACCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGAG 91

Qy   2677 atcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctac 2736
          || |  |||||| |||||||||||| |||||||||||||| | |||||||  | |||||
Db     92 ATAATTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACC-ATCTCTAC 150

Qy   2737 taaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggga 2796
          ||||||||||||||||||  || |||||||| || |||||||||||| | |||||||||
Db    151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy   2797 ggctgaggcatgagaatcacttaaacctgaggaggcagaggttacagtgagccgagattgt 2856
          ||| ||||||  |||||  ||| |||||  || || |||||| ||| |||| | ||||||
Db    211 GGCGGAGGCAGAAGAATTGCTTGAACCTAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGT 270

Qy   2857 gccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaaaaaaaaaa 2915
          ||||||||| |||||||||||  ||||  |||||||||||||||||| | ||||||
Db    271 GCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAT 330

Qy   2915 -------taatcagggcacagtgctctgtgctgtaatcccagctactcgggaggctgag 2967
                 | | |||| |||||||| ||||||||||||| |||| ||| | |||||
Db    331 AATTGTTCTAGCTGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 390

Qy   2968 gtgggtggatcacctgaggtca[illegible] 3027
          | |||||||| |  |||  | |||||||||||||||| | ||  ||| |||  |||
Db    391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGTGGTGAAACCCC 448

Qy   3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          || |||  |||||||||||||||||| |||||||||| |   || ||
Db    449 ATCGCTACAAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCATG 497


RESULT  15
PCT-US95-10202-20
; Sequence 20, Application PC/TUS9510202
;  GENERAL INFORMATION:
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Emi, Mitsuru
;    APPLICANT:  Nakamura, Yusuke
;    APPLICANT:  Durocher, Francine
;    TITLE OF INVENTION:  In Vivo Mutations and Polymorphisms
;    TITLE OF INVENTION:  in the 17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10202
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US 08/409,305
;       FILING DATE:  24-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/348,824
;       FILING DATE:  29-NOV-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08-308,104
;       FILING DATE:  16-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/300,266
;       FILING DATE:  02-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/299,221
;       FILING DATE:  12-AUG-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ihnen, Jeffrey L.
;       REGISTRATION NUMBER:  28,957
;       REFERENCE/DOCKET NUMBER:  24884-109247
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-962-4810
;       TELEFAX:  202-962-8300
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6769 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
PCT-US95-10200-20
```

```
  Query Match             9.1%;  Score 281.4;  DB 5;  Length 6769;
  Best Local Similarity   81.0%;  Pred. No. 3.7e-60;
  Matches  380;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    4;

Qy   2618 gggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcaggcgg 2676
          |||| |||    | ||| | ||||||||| ||| |||| ||||||| ||||||||| |
Db     82 GGGCTGGGCGTAGTAGCTTGCACCTGTAATCCCAGCACTTCGGGAGGCTGAGGCAGGCAG 91

Qy   2677 atcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctac 2736
          || |  |||||| ||||||||||| ||| |||||||||||||||||||| |||||||||
Db     92 ATAATTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACC-CATCTCTAC 150

Qy   2737 taaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttggga 2796
          ||||||||||||||||||  || |||||| | || |||||||||||||||||| ||||
Db    151 TAAAAATACAAAAATTAGCTGGGTGTGGTGGCACGTACCTGTAATCCCAGCTACTCGGGA 210

Qy   2797 ggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagccaagatcgt 2856
          || ||||| |||||  ||| |||||  |  || |||||| ||| ||||||||||||
Db    211 GGCGGAGGCACAAGAATTGCTTGAACCTAGGAGGCGGAGGTTGCAGCGAGCCAAGATCGC 270

Qy   2857 gccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataaaaaa-- 2915
          ||||||||| ||||||||||||   |||| ||||||||||||||||| | ||||||
Db    271 GGCACTGTACTCCAGCCTGGGCGATAGAGTGAGACTCTGTCTCAAAAAAGAAAAAAAAGT 330

Qy   2915 -------taatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgag 2967
                | | |||| |||||||| ||| ||||||||||||||| |||||||| ||
Db    331 AATTGTTCTAGCTGGGCGGCAGTGGCTCTTGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 390

Qy   2968 gtgggtggatcacctgaggtcaggagtttcaagaccagcctggtgaacatggcgaaacccc 3027
          | |||||||| |  ||| |  ||||||| ||||||| |  ||  ||| |||||||||
Db    391 GCGGGTGGATCTC--GAGTCCTAGAGTTCAAGACCAGCCTAGGCAATGTGGTGAAACCCC 448

Qy   3028 gtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          || ||  |||||||||||||||||| ||| |||||| |   || ||
Db    449 ATCGCTACAAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCATG 497
```

```
Search completed: February 10, 2000, 22:54:05
Job time: 1343 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 04:24:46 ; Search time 5888.89 Seconds
(without alignments)
-1586.023 Million cell updates/sec

Title: US-09-174-937-1
Perfect score: 3076
Sequence 1 ccaaggctccgccaagggt. . .....ggcatggtggtgcatgcctg 3076

Scoring table: IDENTITY_NUC

Searched: 921193 seqs, 1518192014 residues

Database : GenEmbl:*

Word size : 0

Number of hits that pass the threshold . 1642386

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1599.6	52.0	179301	41	AC011488	AC011488 Homo sapi
2	348.8	11.3	85654	10	HSP37B26	AL022393 Homo sapi
3	342.4	11.1	178087	40	AC005289	AC005289 Homo sapi
c 4	340.2	11.1	141857	41	AC010532	AC010532 Homo sapi
5	339.6	11.0	42696	11	AC005932	AC005932 Homo sapi
c 6	335.8	10.9	172042	40	AC012670	AC012670 Homo sapi
c 7	333.4	10.8	100000	10	AP000071	AP000071 Homo sapi
8	333.4	10.8	163218	11	AC005837	AC005837 Homo sapi
c 9	333.4	10.8	194502	33	HS1009E24	AL109914 Homo sapi
c 10	333.4	10.8	100364	40	AC007591	AC007591 Homo sapi
11	333.4	10.8	200644	45	AC017113	AC017113 Homo sapi
c 12	332	10.8	183451	40	AC005972	AC005972 Homo sapi
13	331	10.8	185997	30	AC007615	AL117611 Homo sapi
14	331	10.8	156641	40	HUAC002544	AC002544 Homo sapi
15	330.8	10.8	122294	33	AC018931	AL121910 Homo sapi
c 16	330.6	10.7	86769	10	HSAP003626	AP003626 Homo sapi
17	330.4	10.7	112424	32	HS1018A21	AL109868 Homo sapi
18	330.2	10.7	24203	41	AC009101	AC009101 Homo sapi
c 19	329.6	10.7	186881	42	AC008040	AC008040 Homo sapi
20	329.4	10.7	150196	11	HS091126	Z91328 Human Chrom
21	329.4	10.7	142788	31	HS1047L22	AL121844 Homo sapi
c 22	329	10.7	85500	10	HS199M16	AL 22721 Human DNA
23	328.6	10.7	135964	11	AC005914	AC005914 Homo sapi
24	328.2	10.7	192207	40	AC005877	AC005877 Homo sapi
25	327.6	10.7	214238	33	HS1104A21	AL121781 Hum. sapi
26	327.6	10.7	327519	42	AC007621	AC007621 Homo sapi
27	327	10.6	50828	11	HS1191B2	AL122117 Human DNA
28	327	10.6	143376	42	AC012272	AC012272 Homo sapi
29	326.8	10.6	177375	40	AC011462	AC011462 Homo sapi
c 30	326.8	10.6	125109	40	AP000453	AP000453 Homo sapi
31	326.2	10.6	242690	41	AC011501	AC011501 Homo sapi
c 32	326.6	10.6	122948	11	HS391B21	AL109911 Human DNA
c 33	325.6	10.6	163417	40	AC005043	AC005043 Homo sapi
34	325.4	10.6	164079	40	AC005921	AC005921 Homo sapi
c 35	325.4	10.6	150351	42	AC009814	AC009814 Homo sapi
36	325.2	10.6	212731	32	HSCA10957	AL121871 Homo sapi
c 37	325.2	10.6	158062	40	AC005388	AC005388 Homo sapi
c 38	325.2	10.6	99421	40	AP001791	AP001791 Homo sapi
c 39	324.8	10.6	122554	10	AB023049	AB023049 Homo sapi
c 40	324.8	10.6	200001	10	AP000512	AP000512 Homo sapi
c 41	324.8	10.6	99200	11	HS996C27	AL121597 Human DNA
c 42	324.8	10.6	33537	11	HSEA102M1	AL109910 Human DNA
c 43	324.6	10.6	91927	11	AC004771	AC004771 Homo sapi
c 44	324.6	10.6	79516	40	AC004834	AC004834 Homo sapi
45	324.6	10.6	164179	40	AC007227	AC007227 Homo sapi

ALIGNMENTS

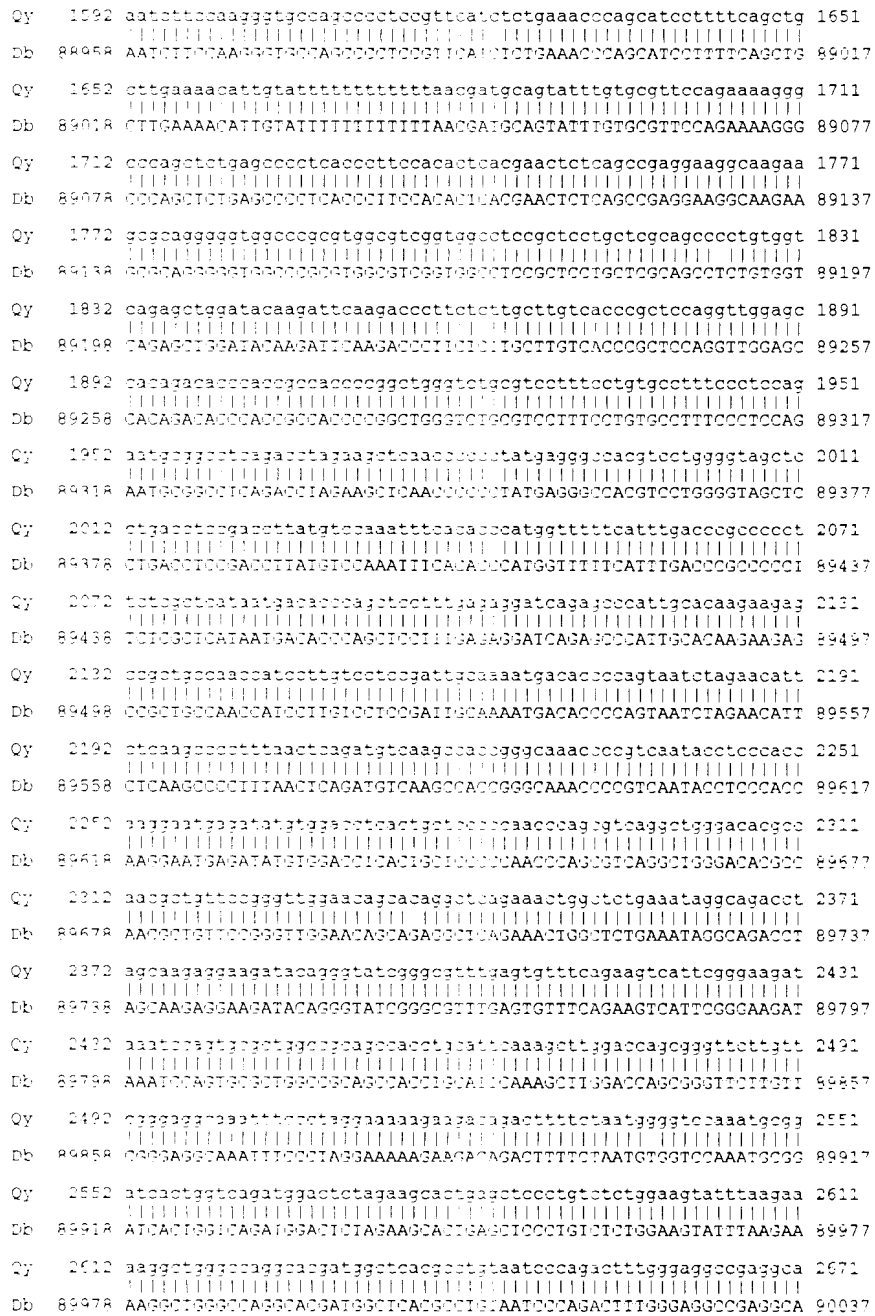
RESULT 1
AC011488
LOCUS AC011488 179301 bp DNA HTG 07-OCT-1999
DEFINITION Homo sapiens chromosome 19 clone CIT978SK_1A1N10, *** SEQUENCING IN PROGRESS ***, 54 unordered pieces.
ACCESSION AC011488
VERSION AC011488.1 GI:6015264
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

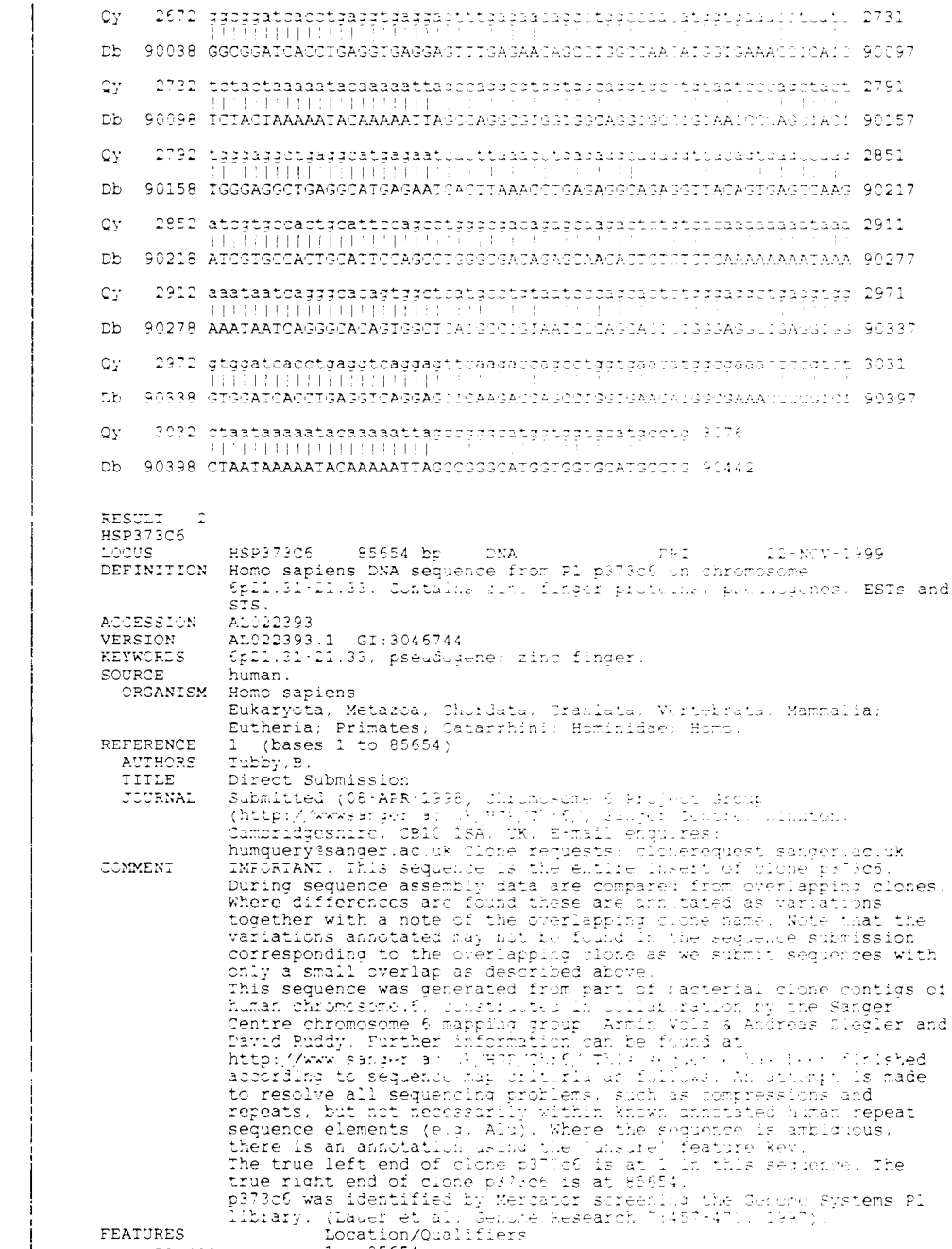
```
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *              gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *              gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *              gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *              gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *              gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *              gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *              gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *              gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *              gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *              gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *              gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *              gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *              gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *              gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *              gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *              gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *              gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *              gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *              gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *              gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *              gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *              gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *              gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *              gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *              gap of unknown length
            *    21389    22058: contig of 670 bp in length
            *              gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *              gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *              gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *              gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *              gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *              gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *              gap of unknown length
            *    27117    27904: contig of 788 bp in length
            *              gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *              gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *              gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *              gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *              gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *              gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *              gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *              gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *              gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *              gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *              gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *              gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *              gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *              gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *              gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *              gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *              gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *              gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *              gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *              gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *              gap of unknown length
            *   112601   179301: contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42599 t      95 others
ORIGIN


  Query Match           52.0%;  Score 1598.6;  DB 41;  Length 179301;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 1601;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    1472 caagcatgtcctgaggcgacgtgagcataacaaaaggacatggaaagaaaattccagatgcc 1531
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   88838 CTAGCATGTCCTGAGGCGACGTGAGCATAACAAAAGGACATGGAAGAAGATTCCAGATGCC 88897

Qy    1532 agaaaacctctgtcagacgcccactgaccccagatctcatcctacttctatcttgatgcc 1591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   88898 AGAAAACCTCTGTCAGACGCCCACTGACCCCAGATCTCATCCTACTTCTATCTTGATGCC 88957
```





RESULT 2
HSP373C6
LOCUS HSP373C6 85654 bp DNA PRI 22-NOV-1999
DEFINITION Homo sapiens DNA sequence from PI p373c6 on chromosome 6p21.31-21.33. Contains 21 finger pseudogenes, pseudogenes, ESTs and STS.
ACCESSION AL022393
VERSION AL022393.1 GI:3046744
KEYWORDS 6p21.31-21.33; pseudogene; zinc finger.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 85654)
AUTHORS Tubby,B.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT IMPORTANT: This sequence is the entire insert of clone p373c6. During sequence assembly data are compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed in collaboration by the Sanger Centre chromosome 6 mapping group (Alan Bruce & Andrea Ciliberto) and David Ruddy. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6/. This project is being finished according to sequence map criteria as follows. No attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'misc_feature' feature key.
The true left end of clone p373c6 is at 1 in this sequence. The true right end of clone p373c6 is at 85654.
p373c6 was identified by Mercator screening the Genome Systems PI library. (Lauer et al. Genome Research 7:487-496, 1997).
FEATURES Location/Qualifiers
source 1..85654
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p21.31-21.33"
                     /clone="XX-P373C6"
     prim_transcript 455. .8814
                     /note="match: multiple ESTs, match: AA580989 H41302
                     AA144413 AA457998"
     gene            455. .7319
                     /gene="p373c6.1"
     CDS             join(455. .1012,5343. .5466,6400. .7319)
                     /gene="p373c6.1"
                     /note="match: AB007897 Q71601: zinc finger protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="p373c6.1"
                     /protein_id="CAA18532.1"
                     /db_xref="GI:3046745"
                     /db_xref="SPTREMBL:O60993"
                     /translation="MAREPPKNAALDAQSAFDQTGLLTVKVEKEEASALTAEVPAPCS
                     PAPGPERSPQPPWGFPYPFAAGPPEALSPLPELCGQWLQPEMHSKEQILELLVLEQFL
                     TILPGNLQSWVPEQHPESGEEVVVLLEYLEPQLDEPAPQVEPTGLVSERCGLFPPEIP
                     DRSEGHSFKVQEPFISFCTVNLLGSKGLLKMEDVALTLTPGWTQLDSSQVNLYRDEKQ
                     ENHSSLVSLGGEIQTKSPDLPPVKKLPEKEHGKICHLPEDIAQIPTHAEAGEQEGRLQ
                     RKQKNAIGSPRHYCHECGKSFAQSSGLTKHPPIHTGEKPYECEDCGKTFIGSSALVIH
                     QRVHTGEKPYECEECGKVFSHSSNLIKHQPTHTGEKPYECDDCGKTFSQSCSLLEHHK
                     IHTGEKPYQCNMCGKAFRENSHLLRHQRIHCDKNVQNPEHGESWESQGRTESQWENTE
                     APVSYKCNECERSFTFNRSLILHQKIHTGEKPYQCDTCGKSFTRTSYLVQHQRSHVGK
                     KILSQ"
     repeat_region   1744. .1876
                     /note="MIR repeat: matches 256. .124 of consensus"
     repeat_region   1949. .2056
                     /note="MIR repeat: matches 3. .112 of consensus"
     repeat_region   2267. .2370
                     /note="AluSp repeat: matches 1  303 of consensus"
     repeat_region   2995. .3129
                     /note="MIR2 repeat: matches 145  4 of consensus"
     repeat_region   3589. .3888
                     /note "AluSx repeat: matches 302. .1 of consensus"
     prim_transcript <5341. .7795
                     /note="match: multiple ESTs, match: N44712 N35344 AA580929
                     M85018 N87411, match: AA227617 F06540 W01293 T08471
                     AA218821; match: N87644 AA333377 R37288"
     repeat_region   7140. .7307
                     /note="2 copies of 94 mer 91 % conserved"
     misc_feature    complement(7766. .8103)
                     /note="match: STS G31316"
     prim_transcript 8264. .8757
                     /note "match: multiple ESTs, match: F02809 AA226879
                     T08470"
     repeat_region   8471. .8772
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   9652. .9677
                     /note="13 copies of 2 mer 92 % conserved"
     repeat_region   12342. .12375
                     /note="17 copies of 2 mer 82 % conserved"
     repeat_region   12398. .12458
                     /note "L1PA15 repeat: matches 844. .904 of consensus"
     repeat_region   13456. .13876
                     /note="MSTA repeat: matches 1. .426 of consensus"
     repeat_region   14248. .14516
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   14607. .14787
                     /note="L1 repeat: matches 3843. .4024 of consensus"
     repeat_region   15512. .15936
                     /note="MER7B repeat: matches 456. .17 of consensus"
     repeat_region   16980. .17326
                     /note="MER4B repeat: matches 205. .539 of consensus"
     repeat_region   17898. .18188
                     /note="AluJo repeat: matches 300. .2 of consensus"
     repeat_region   18496. .18595
                     /note="MER42B repeat: matches 1194  .1300 of consensus"
     repeat_region   18628. .18738
                     /note="MIR repeat: matches 89. .196 of consensus"

     prim_transcript 18958. .19498
                     /note="match: multiple ESTs; match: H10027 AA252169
                     AA412406 AA412407, match: R42157 AA242984"
     misc_feature    <18959. .19309
                     /note="match: STS G23043"
     mRNA            complement(18969  27096)
                     /gene="p373c6.mRNA"
                     /note="match: G62392 zinc finger protein"
                     /evidence=not_experimental
                     /product="p373c6.2.mRNA"
     gene            complement(18969  27096)
                     /gene="p373c6.mRNA"
     gene            complement(19257. .25350)
                     /gene="p373c6.2"
     CDS             complement(join(19257  19873,24599  24745,24991  25350))
                     /gene="p373c6.2"
                     /note="match: G62392 zinc finger protein, match:  15535
                     PRD51 zinc finger protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="p373c6.2"
                     /protein_id="CAA18533.1"
                     /db_xref="GI:3046746"
                     /db_xref="SPTREMBL:O15535"
                     /translation="MNINSKFVLSICVQVPEAWEFLITMKVPAKSHILWLESPLKRSN
                     PLAPEIFPPHFPQLTYLETPGLPEALTRLQELCYQWLRPPVSTKEQILTFLVLEQFLS
                     ILPKELQGWVREHCPESGEEAVILLEDLEPELDEPQHEMVAHRHPQFVLCKEMVPLAE
                     DTPLILQSQPKEPLLICDSAGKCHSIGFTLEVTKTFLPLVLSKQCFPIVEFHQKMFN
                     EQTWEVSQQDPSHLEVGEHKDRIEPQWGNLLGFGQRPPCPRCKSFTQSSSLIRHQRIH
                     TGEPPYECNECGKAFSPSSGLFNHPGIHNIQKPYHCKECGKVFSQSATLIQHQRIHKG
                     EKPYQCSQCSKSYSRPSFLIEHQPSHTGERPHQCIPCGKSFNRHQNLIRHQKIHTVAE
                     LV"
     prim_transcript complement(19549. .20099)
                     /gene="p373c6.2"
                     /note="match: multiple ESTs; match: AA370645 AA47 125
                     AA326745 AA098301"
     repeat_region   20087. .20384
                     /note="AluSx repeat: matches 300  2 of consensus"
     repeat_region   20584. .20658
                     /note="MER44A repeat: matches 333. .260 of consensus"
     repeat_region   20687. .20969
                     /note="AluSg repeat: matches 298  2 of consensus"
     repeat_region   20970. .21425
                     /note="MER44B repeat: matches 442  2 of consensus"
     repeat_region   22279. .22579
                     /note="AluY repeat: matches 1. .301 of consensus"
     repeat_region   22953. .22986
                     /note="MIR2 repeat: matches 107. .140 of consensus"
     repeat_region   23818. .24191
                     /note="MLT1P repeat: matches 2  386 of consensus"
     repeat_region   24199  24340
                     /note="MIR repeat: matches 155. .8 of consensus"
     repeat_region   25621. .25765
                     /note="FRAM repeat: matches 166  24 of consensus"
     repeat_region   25790. .25920
                     /note="MIR repeat: matches 47. .183 of consensus"
     repeat_region   26254. .26353
                     /note="MIR repeat: matches 189. .86 of consensus"
     repeat_region   27832  27879
                     /note="24 copies of 2 mer 81 % conserved"
     repeat_region   27947. .28210
                     /note="AluJo repeat: matches 267. .1 of consensus;
                     incomplete repeat"
     repeat_region   29361  29516
                     /note="MLT2_internal repeat: matches 3880. .4034 of
                     consensus"
     repeat_region   29681. .29988
                     /note="AluYb8 repeat: matches 1. .308 of consensus"
     repeat_region   30585. .30779
                     /note="MLT2_internal repeat: matches 4201. .4401 of
                     consensus"
     repeat_region   30840. .30873
                     /note="17 copies of 2 mer 85 % conserved"
```

```
     repeat_region   31811. .31921
                     /note="MER4B repeat: matches 3. .205 of consensus"
     repeat_region   32005. .32984
                     /note="AluSx repeat: matches 299. .1 of consensus"
     repeat_region   33203. .33309
                     /note="L1ME3A repeat: matches 166. .352 of consensus"
     gene            33801. .34792
                     /gene="p373c6.3"
     CDS             33801. .34792
                     /gene="p373c6.3"
                     /note="match: tob4 D64109; match: Q14106"
                     /codon_start=1

  Query Match             11.3%;  Score 348.8;  DB 10;  Length 85654;
  Best Local Similarity   91.0%;  Pred. No. 1e-69;
  Matches  418;  Conservative    0;  Mismatches   37;  Indels    1;  Gaps    1;

Qy    2562 agatgaactctagaagcagtgagctccctgtctctggaagtatttaagaaaaggctgggc 2621
           |||||||||||||||| | |  || | | |  || | ||  ||   | ||||||||||
Db   57303 ATATGCTCTCTACAAGAAATTCACTTCAAATNTAATGATATNAGTATATTAAAAGTAGGC 57391

Qy    2622 cagtcagtatggctcacgcctgtaatcccagcactttgggaggccgaggcaggcggatca 2680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57392 CAGATGCTGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGAGAGGATCA 57441

Qy    2681 cctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctctactaaa 2740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57442 CTTGAGGCCAGGAGTTCAAGACCAGCCTGACCAACATGGTGAAACTCCATCTCTATTAAA 57501

Qy    2741 aatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggaggct 2800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57502 AATGCAAAAATTAGCCAGGTGTGGTAGTAGGTACCTGTAATCCCAGCTACTCAGGAGGCT 57561

Qy    2801 gaggcatgagaatcacttaaacctgagaggcagaggttacagtgagccaagatcgtgcca 2860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57562 GAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAAATCATGCTA 57621

Qy    2861 ctgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataaaaaaataatca 2920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57622 CTGGACTCTAGCCTGGGTGACAGAGCAAGACTCCATCTAAAAAAAAAAAGTAAATGGGGCC 57681

Qy    2921 gggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtggatcac 2980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57682 AGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGGTGGATCAC 57741

Qy    2981 ctgaggtcaggagttcaagaccagcctggccaacatggcgaaaccccgtctctaataaaa 3040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57742 CTAAGGTCAAGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAA 57801

Qy    3041 atacaaaaattagccgggcatggtggtgcatgcctg 3076
           ||||| |||||||||| | |||| |  ||
Db   57802 ATACAAAACTTAGCCAGGCGTGGTGGCTTACGCCTG 57837


RESULT   3
AC005089
LOCUS       AC005089   178087 bp    DNA             PRI       13-AUG-1999
DEFINITION  Homo sapiens clone RG315H11, complete sequence.
ACCESSION   AC005089
VERSION     AC005089.2  GI:5732140
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178087)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178087)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 178087)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 13, 1999 this sequence version replaced gi:2129.2.
FEATURES             Location/Qualifiers
     source          1. .178087
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG315H11"
BASE COUNT    44105 a   42403 c   43951 g   47727 t
ORIGIN


  Query Match             11.1%;  Score 342.4;  DB 4;  Length 178087;
  Best Local Similarity   97.1%;  Pred. No. 2.9e-68;
  Matches  399;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    2;

Qy    2620 gccaggcacgatggctcacgcctgtaatcccagcactttgggaggccgaggcgggcgat 2679
           |||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   50796 GCCAGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGAT 50855

Qy    2679 cacctgaggtgaggagtttgagaccagcctggccaacatggtgaaacctcatctctacta 2738
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50856 CACTTGAGGCCAGGAGTTCAAGACCAGACTGGCCAACATGGCAAAACCCCATCTCTACTA 50915

Qy    2739 aaaatacaaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgggagg 2798
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50916 AAAATACAAAAATTAGCCAAGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTCGGGAGG 50975

Qy    2799 ctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagccaagatcgtgcc 2858
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50976 CTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGTGCC 51035

Qy    2859 actgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaataaaaaaataatc 2918
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51036 CACTGCATTCCAGCCTGGGCAACAGAGTGAGACTCCGTCAAAAAAAAAAAAAAAAAAAAT 51095

Qy    2919 agggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtggatcac 2978
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   51096 CCGAGCACAGTGGCTCACGCCTGTAATTCCAGCACTTTAGGAGGCTGAGGTGGGTGGATCA 51155

Qy    2979 acctgaggtcaggagttcaagaccagcctggccaacatggcgaaaccccgtctctaataaa 3038
           ||  |||||||||||| | ||
Db   51156 AC--GAGGTCAGGAGTTCGAGACCAGCCTGACCAACATAGCGAAACCCCGTCTCTACTAA 51213

Qy    3039 aaatacaaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||||||||||||| | |  |  |  |||||||||||
Db   51214 AAATACAAAAATTAGCCAGGTGTGGTGGCACATGCCTG 51251


RESULT   4
AC010532/c
LOCUS       AC010532   141857 bp    DNA             HTG       17-SEP-1999
DEFINITION  Homo sapiens chromosome 16 clone PAC 11_21907, *** SEQUENCING IN
            PROGRESS ***, 100 unordered pieces.
ACCESSION   AC010532
VERSION     AC010532.1  GI:5882416
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141857)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141857)
  AUTHORS   DOE Joint Genome Institute.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 100 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      744: contig of 744 bp in length
            *                    gap of unknown length
            *      745     1371: contig of 627 bp in length
            *                    gap of unknown length
            *     1372     2098: contig of 727 bp in length
            *                    gap of unknown length
            *     2099     2812: contig of 714 bp in length
            *                    gap of unknown length
            *     2813     3460: contig of 648 bp in length
            *                    gap of unknown length
            *     3461     4008: contig of 548 bp in length
            *                    gap of unknown length
            *     4009     4882: contig of 874 bp in length
            *                    gap of unknown length
            *     4883     5597: contig of 715 bp in length
            *                    gap of unknown length
            *     5598     6272: contig of 675 bp in length
            *                    gap of unknown length
            *     6273     6993: contig of 721 bp in length
            *                    gap of unknown length
            *     6994     7571: contig of 578 bp in length
            *                    gap of unknown length
            *     7572     8247: contig of 676 bp in length
            *                    gap of unknown length
            *     8248     8370: contig of 123 bp in length
            *                    gap of unknown length
            *     8371     9493: contig of 1123 bp in length
            *                    gap of unknown length
            *     9494    10046: contig of 553 bp in length
            *                    gap of unknown length
            *    10047    11400: contig of 1354 bp in length
            *                    gap of unknown length
            *    11401    12049: contig of 649 bp in length
            *                    gap of unknown length
            *    12050    12626: contig of 577 bp in length
            *                    gap of unknown length
            *    12627    13518: contig of 892 bp in length
            *                    gap of unknown length
            *    13519    14325: contig of 807 bp in length
            *                    gap of unknown length
            *    14326    15013: contig of 688 bp in length
            *                    gap of unknown length
            *    15014    16001: contig of 988 bp in length
            *                    gap of unknown length
            *    16002    16602: contig of 601 bp in length
            *                    gap of unknown length
            *    16603    17537: contig of 935 bp in length
            *                    gap of unknown length
            *    17538    18283: contig of 746 bp in length
            *                    gap of unknown length
            *    18284    18783: contig of 500 bp in length
            *                    gap of unknown length
            *    18784    19778: contig of 995 bp in length
            *                    gap of unknown length
            *    19779    20453: contig of 675 bp in length
            *                    gap of unknown length
            *    20454    21307: contig of 854 bp in length
            *                    gap of unknown length
            *    21308    22076: contig of 769 bp in length
            *                    gap of unknown length
            *    22077    22743: contig of 667 bp in length
            *                    gap of unknown length
            *    22744    23506: contig of 763 bp in length
            *                    gap of unknown length
            *    23507    24399: contig of 893 bp in length
            *                    gap of unknown length
            *    24400    25280: contig of 881 bp in length
            *                    gap of unknown length
            *    25281    26565: contig of 1285 bp in length
            *                    gap of unknown length
            *    26566    27661: contig of 1096 bp in length
            *                    gap of unknown length
            *    27662    27829: contig of 168 bp in length
            *                    gap of unknown length
            *    27830    28926: contig of 1097 bp in length
            *                    gap of unknown length
            *    28927    29670: contig of 744 bp in length
            *                    gap of unknown length
            *    29671    30497: contig of 827 bp in length
            *                    gap of unknown length
            *    30498    32096: contig of 1599 bp in length
            *                    gap of unknown length
            *    32097    32585: contig of 489 bp in length
            *                    gap of unknown length
            *    32586    33411: contig of 826 bp in length
            *                    gap of unknown length
            *    33412    34836: contig of 1425 bp in length
            *                    gap of unknown length
            *    34837    35516: contig of 680 bp in length
            *                    gap of unknown length
            *    35517    36444: contig of 928 bp in length
            *                    gap of unknown length
            *    36445    37331: contig of 887 bp in length
            *                    gap of unknown length
            *    37332    37930: contig of 599 bp in length
            *                    gap of unknown length
            *    37931    38852: contig of 922 bp in length
            *                    gap of unknown length
            *    38853    39711: contig of 859 bp in length
            *                    gap of unknown length
            *    39712    40855: contig of 1144 bp in length
            *                    gap of unknown length
            *    40856    41930: contig of 1075 bp in length
            *                    gap of unknown length
            *    41931    43017: contig of 1087 bp in length
            *                    gap of unknown length
            *    43018    44181: contig of 1164 bp in length
            *                    gap of unknown length
            *    44182    45681: contig of 1500 bp in length
            *                    gap of unknown length
            *    45682    47480: contig of 1799 bp in length
            *                    gap of unknown length
            *    47481    49119: contig of 1639 bp in length
            *                    gap of unknown length
            *    49120    50566: contig of 1447 bp in length
            *                    gap of unknown length
            *    50567    51600: contig of 1034 bp in length
            *                    gap of unknown length
            *    51601    52991: contig of 1391 bp in length
            *                    gap of unknown length
            *    52992    54007: contig of 1016 bp in length
            *                    gap of unknown length
            *    54008    55406: contig of 1399 bp in length
            *                    gap of unknown length
            *    55407    56899: contig of 1493 bp in length
            *                    gap of unknown length
            *    56900    59160: contig of 2261 bp in length
            *                    gap of unknown length
            *    59161    61384: contig of 2224 bp in length
            *                    gap of unknown length
            *    61385    63339: contig of 1955 bp in length
            *                    gap of unknown length
            *    63340    64631: contig of 1292 bp in length
            *                    gap of unknown length
```

```
            *    64033    66592: contig of 1961 bp in length
            *                   gap of unknown length
            *    66593    67767: contig of 1175 bp in length
            *                   gap of unknown length
            *    67768    69280: contig of 1513 bp in length
            *                   gap of unknown length
            *    69281    70471: contig of 1191 bp in length
            *                   gap of unknown length
            *    70472    72330: contig of 1859 bp in length
            *                   gap of unknown length
            *    72331    73776: contig of 1446 bp in length
            *                   gap of unknown length
            *    73777    75448: contig of 1672 bp in length
            *                   gap of unknown length
            *    75449    76908: contig of 1460 bp in length
            *                   gap of unknown length
            *    76909    78869: contig of 1961 bp in length
            *                   gap of unknown length
            *    78870    80157: contig of 1288 bp in length
            *                   gap of unknown length
            *    80158    82113: contig of 1956 bp in length
            *                   gap of unknown length
            *    82114    84746: contig of 2633 bp in length
            *                   gap of unknown length
            *    84747    86466: contig of 1720 bp in length
            *                   gap of unknown length
            *    86467    87876: contig of 1410 bp in length
            *                   gap of unknown length
            *    87877    89436: contig of 1560 bp in length
            *                   gap of unknown length
            *    89437    91590: contig of 2154 bp in length
            *                   gap of unknown length
            *    91591    93787: contig of 2197 bp in length
            *                   gap of unknown length
            *    93788    96557: contig of 2770 bp in length
            *                   gap of unknown length
            *    96558    98457: contig of 1900 bp in length
            *                   gap of unknown length

  Query Match           11.1%;  Score 340.2;  DB 41;  Length 141857;
  Best Local Similarity   84.2%;  Pred. No. 9.1e-68;
  Matches  395;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy   2609 gaaaaggctgggccaggcacggatggctcacgcctgtaatcccagactttgggaggccgag 2668
          | | ||    ||| | |||  |||||||| ||||||||    ||||||||||||||||||
Db 123186 GTACAGAGGAGGTTGGAGACAGTCGGCTCATCTTGTAATCGAGCACTTTGGGAGGCCGAG 123427

Qy   2669 gcaggcggatcacctgaggtcaggagttcgagaccagcctgaccaacatggtgaaaccctc 2728
          | |||||||||| ||||||  |||||||| | | ||||||||||||||||||||||||| |
Db 123428 GTAGGCGGATCACTTGAGGTCAGGAGTTTAAGACCAGCCTGGCCAACATGGTGAAACCCC 123367

Qy   2729 atctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagct 2788
          || |||| |||||||||| |||||||||| |||||||| | ||||||||| ||||||||||
Db 123366 ATTTCTATTAAAAATACAAACATTAGCCAGGTTTGGTGGCAGGCGCCTGTAATCCCAGCT 123307

Qy   2789 acttgggaggctgaggcatgagaatcacttgaacctgaggaggcagaggttacagtgagcc 2848
          ||| |||||||||||| ||||||| ||| | || | ||||| |||||| ||||||||||
Db 123306 ACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCC 123247

Qy   2849 aagatggtgccactgcactccagcctgggcgacagagcgaagactctgtctcaaaaaaaa 2907
            ||| | ||||||||  | ||||||||| | |||||  ||||||||||||| ||||| |
Db 123246 TCGATTGTACCATCGCACTGCAGCCTGGGCGACAGAGTGAGACTCTGTCTCACAAAAACA 123187

Qy   2908 taaaaaataatcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgag 2967
          ||| || | | |||||| |||||||||| ||||||||||||||||| |||||||| |||
Db 123186 AAAACAACAGGGTGGCACGGTGGTCATACCTGTAATCCCAGCACTTTGGGAGGCCGAG 123127

Qy   2968 gtgggtggatcacctgaggtcaggagttcgagaccagcctggtgaacatggcgaaaccccg 3027
          |||| |||||||| |||||| |||| || |||||||||  ||||||  |||||||
Db 123126 GTGGGCGGATCACCTGAGGTCAGATGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCT 123067

Qy   3028 gtctctaataaaaatacaaaaattagccgggcgtggtggtgcatgcctg 3076
          ||||||| ||||||| ||| ||| |||| ||  |||||||| ||||||
```

```
Db 123066 GTCTCTACTAAAAATGCAACAATTAGCCGGGCGTGGTGGTGCATGCCTG 123018


RESULT   5
AC005932
LOCUS       AC005932    41696 bp    DNA             PRI       15-NOV-1998
DEFINITION  Homo sapiens chromosome 19, cosmid R31317, complete sequence.
ACCESSION   AC005932
VERSION     AC005932.1  GI:3845350
KEYWORDS    HTG
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41696)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Wyatt,McKinley,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Olsa,S., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Carnes,J.,
            Danganan,L., Erler,A., Christensen,M., Gebregeorgis,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Goefield,J., Duarte,S., Lucas,S., Bruce,D., Thodes,P., Quan,G.,
            Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a ~1 Mb region containing the MEF2B gene in
            19p12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41696)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            R31317 overlaps cosmid R27090 (AC002487) to the left from bases 1
            to 1,849 of this accession, and overlaps cosmid R27918 (AC004447)
            to the right from bases 32,807 to 41,696. Additional chromosome 19
            map and sequence information are available at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1..41696
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R31317"
                     /chromosome="19"
                     /map="19p12 between UBA52 and D19S451"
                     /cell_line="5HL2-B"
                     /clone_lib="LL19NC03 R chromosome 19 cosmid library"
                     /note="LL19NC03 cosmid library constructed at LLNL from
                     flow-sorted chromosomes from hybrid 5HL2 B, which carries
                     chromosome 19 as its only human chromosome."
     repeat_region   complement(47..169)
                     /rpt_family="AluJo/FLAM"
     repeat_region   793..823
                     /rpt_family="MIR"
     repeat_region   1108..1213
                     /rpt_family="FLAM_C"
     repeat_region   1297..1598
                     /rpt_family="AluSx"
     repeat_region   1599..1630
                     /rpt_family="(GGAA)n"
     repeat_region   complement(1678..1977)
                     /rpt_family="AluSc"
     repeat_region   2461..2610
                     /rpt_family="MIR"
     repeat_region   complement(2618..2918)
                     /rpt_family="AluSx"
     repeat_region   complement(3041..3247)
                     /rpt_family="AluJb"
     misc_feature    complement(4137..4216)
                     /note="predicted exon: program: sorf12exons_human_1.3,
                     frame: 2, quality: good, score: 66.1000"
     repeat_region   4408..4709
                     /rpt_family="AluSq"
     repeat_region   complement(4759..5055)
```

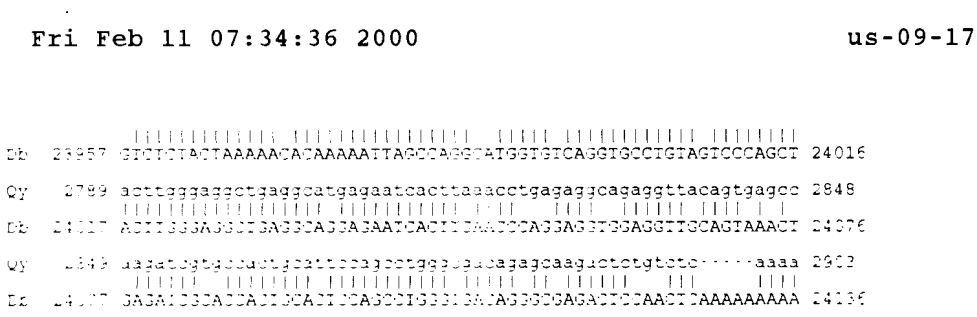
```
                     /rpt_family="AluSx"
     repeat_region   complement(5239. .5481)
                     /rpt_family="LINE2"
     repeat_region   complement(6476. .6778)
                     /rpt_family="AluY"
     repeat_region   complement(6784. .7078)
                     /rpt_family="AluJo"
     repeat_region   complement(7082. .7136)
                     /rpt_family="MIR"
     repeat_region   7151. .7282
                     /rpt_family="AluSq/x"
     repeat_region   7287. .7587
                     /rpt_family="AluY"
     repeat_region   7603. .7907
                     /rpt_family="AluSx"
     repeat_region   7908. .8206
                     /rpt_family="AluY"
     repeat_region   8217. .8518
                     /rpt_family="AluSx"
     repeat_region   8690. .8732
                     /rpt_family="MIR"
     repeat_region   complement(8935. .9174)
                     /rpt_family="LINE2"
     repeat_region   complement(9176. .9359)
                     /rpt_family="AluJb"
     repeat_region   9371. .9667
                     /rpt_family="AluSx"
     repeat_region   9677. .9722
                     /rpt_family="Alu"
     repeat_region   9954. .10255
                     /rpt_family="AluSx"
     repeat_region   10263. .10437
                     /rpt_family="AluJo"
     repeat_region   complement(11193. .11255)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(11261. .11879)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(11736. .11859)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(11924. .12186)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(12216. .12328)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(12331. .12401)
                     /rpt_family="(TGGA)n"
     repeat_region   12819. .12935
                     /rpt_family="MIR"
     repeat_region   13085. .13385
                     /rpt_family="AluJb"
     repeat_region   complement(13484. .13651)
                     /rpt_family="AluSx"
     repeat_region   complement(13652. .13956)
                     /rpt_family="AluSg"
     repeat_region   complement(13957. .14091)
                     /rpt_family="AluSx"
     repeat_region   14326. .14370
                     /rpt_family="LINE2"
     repeat_region   complement(14437. .14464)
                     /rpt_family="L1M7_A"
     repeat_region   complement(14655. .14953)
                     /rpt_family="AluSx"
     repeat_region   complement(14986. .15308)
                     /rpt_family="7SLRNA"
     repeat_region   15483. .15608
                     /rpt_family="FLAM_C"
     repeat_region   complement(15616. .15894)
                     /rpt_family="AluSx"
     repeat_region   complement(16128. .16426)
                     /rpt_family="AluSx"
     repeat_region   16465. .16767
                     /rpt_family="AluJo"
     repeat_region   16777. .17059
                     /rpt_family="AluJb"
     repeat_region   17091. .17389
                     /rpt_family="AluSx"
     repeat_region   complement(17536. .17604)
                     /rpt_family="AluJ/FLAM"
     repeat_region   complement(17612. .17644)
                     /rpt_family="MIR"
     repeat_region   complement(17719. .17888)
                     /rpt_family="AluSq/x"
     repeat_region   complement(17889. .18192)
                     /rpt_family="AluY"
     repeat_region   complement(18193. .18491)
                     /rpt_family="AluSx"
     repeat_region   19366. .19490
                     /rpt_family="LINE2"
     repeat_region   19658. .19956
                     /rpt_family="AluSx"
     repeat_region   20017. .20317
                     /rpt_family="AluSq"
     repeat_region   20452. .20752
                     /rpt_family="AluSx"
     repeat_region   20771. .21070
                     /rpt_family="AluJo"
     repeat_region   21073. .21369
                     /rpt_family="AluSq"
     repeat_region   21370. .21446
                     /rpt_family="(CAA)n"
     repeat_region   21469. .21576
                     /rpt_family="AluY"
     repeat_region   21544. .21637
                     /rpt_family="LIR9"
     repeat_region   complement(21644. .21943)
                     /rpt_family="AluSq"
     repeat_region   complement(21944. .22232)
                     /rpt_family="AluSq"
     repeat_region   complement(22255. .22556)
                     /rpt_family="AluSq"
     repeat_region   22567. .22744
                     /rpt_family="LIR9"
     repeat_region   22760. .23027
                     /rpt_family="AluSq"
     repeat_region   23031. .23091
                     /rpt_family="LIR9"
     repeat_region   23094. .23393
                     /rpt_family="AluSq"
     repeat_region   23401. .23496
                     /rpt_family="LIR9"
     repeat_region   23479. .23740
                     /rpt_family="LIR9"
     repeat_region   23754. .23845
                     /rpt_family="AluJb"
     repeat_region   23846. .24147
                     /rpt_family="AluSx"
     repeat_region   24151. .24470
                     /rpt_family="AluSx"
     repeat_region   24851. .24894
                     /rpt_family="(TAAA)n"
     repeat_region   25113. .25161
                     /rpt_family="MIR"
     repeat_region   complement(25169. .25196,

  Query Match             11.0%;  Score 338.6;  DB 11;  Length 41696;
  Best Local Similarity   85.2%;  Pred. No. 2.1e-67;
  Matches  403;  Conservative    0;  Mismatches   64;  Indels    6;  Gaps    2;

Qy    2618 aaaaggctggaccaggcacgatggctcacgcctataatcccag actttgggaggc 2677
           | |||   |||| ||||| | |||||||| ||||||||||||| || |||| |||
Db   23837 ATAAGCACAGGGCTAGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC 23896

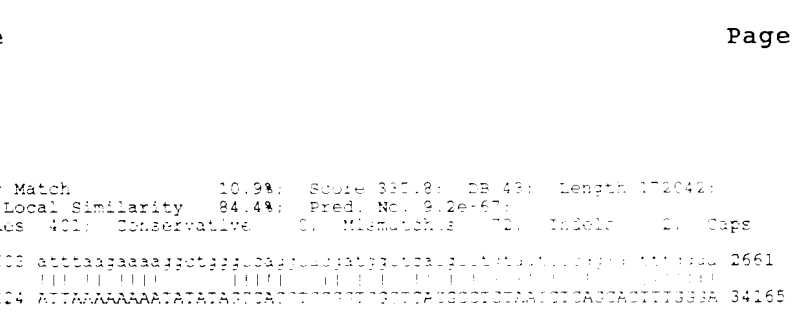
Qy    2678 gcaggcggatcacctgaggtgaggagttcgagaccagcctgaccaacatggtgaaacct 2737
           || ||| |||||||||||||| |||||||| |||| |||||||||||| |||| || ||
Db   23897 GCGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCC 23956

Qy    2738 atctctactaaaaatacaaaaattagccaggcatggtggcgggcacctgtaatcccagc 2797
```



RESULT 6
AC012676/c
LOCUS AC012676 172042 bp DNA HTG 03-NOV-1999
DEFINITION Homo sapiens chromosome 16 clone RPCI-11_295D4, *** SEQUENCING IN PROGRESS ***, 6 ordered pieces.
ACCESSION AC012676
VERSION AC012676.1 GI:6023624
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172042)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172042)
AUTHORS DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 1889: contig of 1889 bp in length
* gap of unknown length
* 1890 6430: contig of 4541 bp in length
* gap of unknown length
* 6431 15555: contig of 9125 bp in length
* gap of unknown length
* 15556 17583: contig of 2028 bp in length
* gap of unknown length
* 17584 59849: contig of 42266 bp in length
* gap of unknown length
* 59850 172042: contig of 112193 bp in length.
FEATURES Location/Qualifiers
source 1..172042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RPCI-11_295D4"
BASE COUNT 39211 a 47582 c 47356 g 38879 t 15 others

ORIGIN



RESULT 7
AP000070/c
LOCUS AP000070 100000 bp DNA PRI 20-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 8p11.2, sequencing gap region, section 6/19, complete sequence.
ACCESSION AP000070
VERSION AP000070.1 GI:4579991
KEYWORDS HTG.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isomura,M., Ikegawa,S., Kinjo,T. and Nakamura,Y.
TITLE DNA sequence analysis of a 1.9-Mb region on chromosome 8p11.2
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 100000)
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,T.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1999) [illegible] Mika Hirakawa, Japan Science and Technology Corporation (JST), Advanced Database Department, 3-3, Yumian-cho, Chiyoda-ku, Tokyo 102-0028, Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491, Fax:81-3-5214-8470)
COMMENT >>text_table: 0000140177.
FEATURES Location/Qualifiers
source 1..100000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p11.2"
STS 17263..17588
/standard_name="D8S449"

```
                     /note="UT5368,The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:198685"
BASE COUNT     26805 a  21045 c  22390 g  28738 t      32 others
ORIGIN


  Query Match             10.8%;  Score 333.4;  DB 10;  Length 100000;
  Best Local Similarity   84.1%;  Pred. No. 3.2e-66;
  Matches  401;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    2;

Qy    2605 ttaggaaaaggctgggccaggcacgatgactcacgcctgtaatcccag-actttgggagg 2663
           ||   |||| |  |||||||   | || ||| | |||| | |||||| | ||||||||
Db    8505 TTTTAAAAATGAAGGGCCAGGAGTGGTGTCTCACACCTGCATTCCCAGCATTTTGGGAGC 8446

Qy    2664 cgaggcaggcggatcacctgaggtgaggtagtttgagaacagcctggccaacatggtgaa 2723
           || |||||||| |||||||||||||| || |||||||| ||||||||||||||||||||
Db    8445 GCTAGGCAGGCAGATCACCTGAGGTCAGTAGTTTGAGACCAGCCTGGCCAACATGGTGAA 8386

Qy    2724 acctcgtctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcc 2783
           ||||||||||||||||||||||||||||||||| |||| |||||||| |  ||||||||||
Db    8385 ACCTCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAAGCACCTGTAATCC 8326

Qy    2784 cagctacttgggaggctgaggcatgagaatcacttaaacctgaggaggcagaggttacagt 2843
           |||||| |||||||||| ||| |||| |  ||| |||   ||||| |||||| |||
Db    8325 CAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTTATAGT 8266

Qy    2844 gagccgagatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaa 2903
           ||  | ||| ||||||||||| ||||| || || |||||||  |||||| | |||||||
Db    8265 GAGATCAAAATTGTGCCATTGCACTCCAGCCTGGGAGACAGAGTGAGACTCCGCCTCAAAA 8206

Qy    2904 aaaataaaaaaataatca----ggcacagtgactcatgcctgtaatcccagcactctggg 2959
           |||| |||| |  |    |||||| ||  ||| |||||||||||||||||  |||
Db    8205 AAAAAAAAAAAAAATGAAGAATGGGCACAGTGGCTCACGCCTGTAATCCCAGCACTTAGGG 8146

Qy    2960 aggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggc 3019
           |||| |||| ||||||||||||||||| |||||||| ||| ||||||||  ||||||||
Db    8145 AGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCGAGATCAGCCTGGCCAACATGGC 8086

Qy    3020 gaaaccccgtctctactaaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           |||||  |||||||  |||||||||||| ||||||| ||||||||||  ||  ||||
Db    8085 CAAACCTTGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTG 8029


RESULT   8
AC005837
LOCUS       AC005837   163218 bp    DNA             PRI       06-NOV-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.318_A_15, complete sequence.
ACCESSION   AC005837
VERSION     AC005837.1  GI:3849820
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.318_A_15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,P., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,T., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 163218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,P., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,T., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 6, 1998 this sequence version replaced gi:3845403.
            All repeats were identified using RepeatMasker: Smit, A. F. A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1..163218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.318_A_15"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   1..233
                     /rpt_family="AluSg/x"
     repeat_region   246..447
                     /rpt_family="L1MB8"
     repeat_region   448..813
                     /rpt_family="LTR22"
     repeat_region   complement(814..1034)
                     /rpt_family="AluY"
     repeat_region   1035..1164
                     /rpt_family="L1R22"
     repeat_region   1165..1553
                     /rpt_family="L1MB8"
     repeat_region   complement(1554..1849)
                     /rpt_family="AluY"
     repeat_region   1850..1871
                     /rpt_family="L1MB8"
     repeat_region   complement(2022..2285)
                     /rpt_family="L1MC/D"
     repeat_region   complement(2298..2566)
                     /rpt_family="AluSq"
     repeat_region   complement(2567..2660)
                     /rpt_family="L1MC/D"
     repeat_region   2661..2971
                     /rpt_family="AluSq"
     repeat_region   complement(2972..2998)
                     /rpt_family="L1MC/D"
     repeat_region   2999..3192
                     /rpt_family="Tigger2a"
     repeat_region   complement(3193..3504)
                     /rpt_family="AluSq"
     repeat_region   3505..3587
                     /rpt_family="Tigger2a"
     repeat_region   3588..3733
                     /rpt_family="AluSx"
     repeat_region   3734..4039
                     /rpt_family="AluY"
     repeat_region   4040..4210
```

```
                     /rpt_family="AluSx"
     repeat_region   complement(4215. .4527)
                     /rpt_family="AluSq"
     repeat_region   4528   4697
                     /rpt_family="Tigger2a"
     repeat_region   complement(4698. .4877)
                     /rpt_family="L1MC/D"
     repeat_region   complement(4858. .4887)
                     /rpt_family="L1MC/D"
     repeat_region   complement(4888. .5192)
                     /rpt_family="AluJb"
     repeat_region   complement(5193. .5327)
                     /rpt_family="L1MC/D"
     repeat_region   complement(5464. .5761)
                     /rpt_family="AluSq1"
     repeat_region   complement(6610. .6911)
                     /rpt_family="AluSx"
     repeat_region   7067. .7367
                     /rpt_family="AluSx"
     repeat_region   7369. .7677
                     /rpt_family="AluSp"
     repeat_region   7678. .7793
                     /rpt_family="(TA)n"
     repeat_region   complement(8263. .8554)
                     /rpt_family="AluSq"
     repeat_region   8555. .8730
                     /rpt_family="MIR"
     repeat_region   complement(8791. .9089)
                     /rpt_family="AluSx"
     repeat_region   complement(9775   10099)
                     /rpt_family="AluSq1"
     repeat_region   complement(10477. .10578)
                     /rpt_family="U6"
     repeat_region   10795. .11092
                     /rpt_family="AluJb"
     repeat_region   11537. .11702
                     /rpt_family="MIR"
     repeat_region   11703. .11762
                     /rpt_family="AT_rich"
     repeat_region   complement(11765. .12074)
                     /rpt_family="AluJb"
     repeat_region   complement(12087. .12381)
                     /rpt_family="AluYb8"
     repeat_region   12411. .12682
                     /rpt_family="AluSx"
     repeat_region   12835. .13125
                     /rpt_family="AluY"
     repeat_region   13376. .13583
                     /rpt_family="L2"
     repeat_region   complement(13584. .13880)
                     /rpt_family="AluYb8"
     repeat_region   13881. .13894
                     /rpt_family="L2"
     repeat_region   complement(13932. .14022)
                     /rpt_family="AluJb"
     repeat_region   14279. .14433
                     /rpt_family="L1MP8"
     repeat_region   14444. .14515
                     /rpt_family="L2"
     repeat_region   14554. .14676
                     /rpt_family="purine-rich"
     repeat_region   complement(14793. .15084)
                     /rpt_family="AluJb"
     repeat_region   15318. .15573
                     /rpt_family="L2"
     repeat_region   15859. .16150
                     /rpt_family="AluSq"
     repeat_region   complement(16151   16257)
                     /rpt_family="AluJb"
     repeat_region   complement(16264. .16410)
                     /rpt_family="Tigger1"
     repeat_region   16411. .16576
                     /rpt_family="AluSq/x"
     repeat_region   complement(16579. .17524)
                     /rpt_family="Tigger1"
     repeat_region   17525. .17644
                     /rpt_family="TAM"
     repeat_region   17648. .17685
                     /rpt_family="(TAAA)n"
     repeat_region   complement(17686. .18134)
                     /rpt_family="Tigger1"
     repeat_region   complement(18722. .18800)
                     /rpt_family="MIR"
     repeat_region   19576. .19878
                     /rpt_family="AluSp"
     repeat_region   20126. .20165
                     /rpt_family="AT_rich"
     repeat_region   20687. .20988
                     /rpt_family="AluSq"
     repeat_region   complement(21696. .21998)
                     /rpt_family="AluJb"
     repeat_region   complement(22789. .23101)
                     /rpt_family="MIR"
     repeat_region   complement(23290. .23605)
                     /rpt_family="AluY"
     repeat_region   23825. .23921
                     /rpt_family="L1MA8"
     repeat_region   complement(24154. .24223)
                     /rpt_family="MIR"
     repeat_region   complement(24171. .24234)

  Query Match             10.9%;  Score 355.4;  DB 11;  Length 169218;
  Best Local Similarity   95.2%;  Pred. No. 3.2e-66;
  Matches  395;  Conservative    0;  Mismatches   16;  Indels    2;  Gaps    2;

Qy   2616 ctgggccaggcacgatggctcacgcctgtaatcccagcactttgggaggctgaggcagg 2674
           |||||||||   |||||||||||| |||||| |||| |||||||||||||||||||
Db   7064 CCGAGCCAGGTGTGGTGGCTCACGCCTATAATCCCAGCACTCTGGGAGGCCGAGGCGGG 7123

Qy   2675 ggatcacctgaggtcaggagttcgagaccagcctggccaacatggtgaaacctcatctct 2734
           |||||| |||||| ||||||||||||| ||| || |||| |||||||||||| ||||
Db   7124 GGATCACTTGAGGTCAGGAGTTCGAGACAGCCTGGCCAACATGGTGAAACCCCGTCTCT 7183

Qy   2735 act-aaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctactta 2793
           ||| |||||| |||||||||| |||||| || |||| |||||||||||| |||||||
Db   7184 ACTAAAAAATGCAAAAATTAGCCGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCA 7243

Qy   2794 ggaggctgaggcatgagaatcacttgaacctgggaggcagaggttgcagtgagccgagatt 2853
           |||||||||||| |||||||| |||| |||||||| |||| |||| |||| |||| ||
Db   7244 GTAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGAT 7303

Qy   2854 cgtgccactgcattccagcctgggcgacagagcgagactccgtc tcaaaaaaaaaaaaaa 2913
           |||||| |||| |||||||||||| |||||||||||||||||| |||||||||||||||
Db   7304 GGTGCCAGTGCACTCCAGCCTGGGCAACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAA 7363

Qy   2914 ataatcagggcacagtggctcatgcctgtaatcccagcactttgggaggctgaggtgggt 2973
           | |||| ||| |||||||| | | | | | | | | | | | | | | | | | | |
Db   7364 AAAATCTGGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGT 7423

Qy   2974 ggatcacctgaggtcaggagttcaagaccagcctggccaacatggcgaaaccccgtctct 3033
           |||||||||||||||| |||| | | | | | | | | | | | | | | | | | | |
Db   7424 GGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTAACCAACGTGGAGAAACCCCATCTCT 7483

Qy   3034 aataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | |||||||||||||||| | | | | | | | | | | | |
Db   7484 ACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCACATGCCTG 7526


RESULT   9
HS1009E24/c
LOCUS       HS1009E24  194502 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 20 clone RP5-1009E24, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL109804
VERSION     AL109804.15  GI:6522703
KEYWORDS    HTG; HTGS_PHASE1.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194502)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 3, 1999 this sequence version replaced gi:6469149.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00511  Length: 1294bp
            Contig_ID: 00965  Length: 1694bp
            Contig_ID: 01480  Length: 1990bp
            Contig_ID: 02471  Length: 49345bp
            Contig_ID: 02668  Length: 4355bp
            Contig_ID: 02812  Length: 1474bp
            Contig_ID: 03132  Length: 11026bp
            Contig_ID: 03183  Length: 2271bp
            Contig_ID: 03185  Length: 51182bp
            Contig_ID: 03383  Length: 18802bp
            Contig_ID: 03987  Length: 22594bp.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..194502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-1009E24"
                     /clone_lib="RPCI-5"
BASE COUNT    43780 a  49145 c  49467 g  44105 t   8005 others
ORIGIN
```

```
  Query Match             10.8%;  Score 333.4;  DB 33;  Length 194502;
  Best Local Similarity   95.0%;  Pred. No. 3.2e-66;
  Matches  397;  Conservative    0;  Mismatches   16;  Indels    4;  Gaps    2;

Qy    2614 ggctgggccaggcagcatgatggctcacgcctgtaatcccag-actttgggaggccgaggcag 2672
           || | ||| ||| | |  |||||| |||| ||||||| |||||||| ||| |||| |
Db   18570 GGGGGTGCCGGGCGCTGCGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGG 18511

Qy    2673 gcggatcacctgaggtgaggagttttgagaacagcctggccaacatggtgaaacctcatct 2732
           | |||||||| |||| |||||||| | | |||||||||||| |||||| ||||
Db   18510 GTAGATCACCTGAGGTCAGGAGTTTGGGACCAGCCTGGCCAACATGGCGAAACCCCATCT 18451

Qy    2733 ctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctactt 2792
           ||| | |||||||||||| ||| |||| || ||   |  ||||||||||||||||||
Db   18450 CTACTAAAAATACAAAAATCAGCTAGGCATGGTGGTGCACCTGTAATCCCAGCTACTC 18391

Qy    2793 gggaggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtgagccaaga 2852
           | |||||||||| |||||| ||| | ||  ||||| |||||| ||||||||||||
Db   18390 AGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGA 18331

Qy    2853 tcgtgccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaataaaa 2912
           || |||||||| |||||||| || ||  | |||||||||||||||||| ||||
Db   18330 TCCCGTCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAA 18271

Qy    2913 aataa---tcagggcacaagtggctcatgcctgtaatcccagcactctgggaggctgaggt 2969
           |||   | |||  |||||||| | ||| |||||| |||||| ||||||||||||
Db   18270 AAAAAAAGGCCGGGTGAGGTGGCTCACGCCTGTAATCACAGCACTTTGGGAGGCTGAGGC 18211

Qy    2970 gggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccgt 3029
           | | | | | | | | | |||||||| |  |||||||  ||||||| ||||||| ||
```

```
Db   18210 GGGCGGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGGCCAACATGGAGAAACCCTGT 18151

Qy    3030 ctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           ||||| ||||||    |||||||||| ||| |||||| |||||||||
Db   18150 CTCTACTAAAAAATACAAAATTAGCCAGGCGTGGTGGCGCATGCCTG 18104


RESULT  10
AC007590/c
LOCUS       AC007590   100364 bp    DNA             PRI       19-MAY-1999
DEFINITION  Homo sapiens BAC 161A6, complete sequence.
ACCESSION   AC007590
VERSION     AC007590.1  GI:4874262
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100364)
  AUTHORS   Huang,E.N., Parnell,L.D., Habermann,K., Vil,M.D., Rodrigues,M.,
            Matero,A., Shah,R., Schutz,K., Shekher,M., See,L.H., Spiegel,L.A.,
            Swaby,I., Nascimento,L., O'Shaughnessy,A., Preston,R.R., de la
            Bastide,M., Dedhia,N.N. and McCombie,W.R.
  TITLE     Homo sapiens BAC 161A6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 100364)
  AUTHORS   Huang,E.N., Parnell,L.D., Habermann,K., Vil,M.D., Rodrigues,M.,
            Matero,A., Shah,R., Schutz,K., Shekher,M., See,L.H., Spiegel,L.A.,
            Swaby,I., Nascimento,L., O'Shaughnessy,A., Preston,R.R., de la
            Bastide,M., Dedhia,N.N. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1999) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724
FEATURES             Location/Qualifiers
     source          1..100364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="BAC 161A6"
     misc_feature    93924..93939
                     /note="in some subclone traces the A at position 93932 is
                     called as a G and sequences of PCR products have a double
                     peak; a restriction digest gives no indication of the
                     occurrence of a repeat unit"
BASE COUNT    27307 a  28999 c  22609 g  29450 t
ORIGIN
```

```
  Query Match             10.8%;  Score 333.4;  DB 40;  Length 100364;
  Best Local Similarity   84.1%;  Pred. No. 3.2e-66;
  Matches  401;  Conservative    0;  Mismatches   71;  Indels    5;  Gaps    2;

Qy    2605 ttaagaaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggagg 2663
           ||   |||| |  |||||||   | || |||| |||| | || ||| |||| | ||
Db   95788 TTTAAAAATGAAGGGCCAGGAGTGGTGTCTCACACCTGCAATTCCCAGCATTTGGGAGG 95729

Qy    2664 ccgaggcaggcggatcacctgaggtgaggagttttgagaacagcctggccaacatggtgaa 2723
           || |||||||| ||||||||||||| || ||||||| |||||||| | | |||| |
Db   95728 CCTAGGCAGGCAGATCACCTGAGGTCAGTAGTTTGAGACCAGCCTGGCCAACATGGTGAA 95669

Qy    2724 acctcatctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcc 2783
           |||||||||||||||||||||||||||||||||||||| ||||||| | |||| ||||
Db   95668 ACCTCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAAGCACCTGTAATCC 95609

Qy    2784 cagctacttgggaggctgaggcatgagaatcacttaaacctgaaaggcagaggttacagt 2843
           |||||||||||||||||| ||| ||||| ||| ||  | |||| |||| |||| ||||
Db   95608 CAGCTACTTGGGAGGCTGAAGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTTACAGT 95549

Qy    2844 gagccaagatcgtgccactgcattccagcctgggcgacagagcgagactctgtctcaaaa 2903
           |||   | || |||||||||| |||||| |||| |||.||.   || || | |
Db   95548 GAGATGAAATTGTGCCACTGCACTCCAGCCTGGGAGACAGAGTGAGACTCTGTCTCAAAA 95489

Qy    2904 aaaataaaaaataatca----gggcacaagtggctcatgcctgtaatcccagcactctgaa 2959
```

```
Db  95488 AAAAAAAAAAAAAATGAAGAATGGGCACACTTGGCTCACGCCTGTAATCCCAGCACTTAGGG 95429

Qy   2960 aggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggc 3019
Db  95428 AGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGC 95369

Qy   3020 gaaaccccgtctctactaaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
Db  95368 CAAACCTCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTG 95312
```

RESULT 11
AC017113
LOCUS AC017113 200644 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone RP11-574H7 *** SEQUENCING IN PROGRESS ***, 29 unordered pieces.
ACCESSION AC017113
VERSION AC017113.1 GI:6554026
KEYWORDS HTG; HTGS_PHASE1
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 200644)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 200644)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St Louis, MO 63108, USA
COMMENT Center project name: H_NH0574H07.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	2041: contig of 2041 bp in length
*	2042	2059: gap of unknown length
*	2060	4340: contig of 2281 bp in length
*	4341	4358: gap of unknown length
*	4359	7715: contig of 3357 bp in length
*	7716	7733: gap of unknown length
*	7734	9874: contig of 2141 bp in length
*	9875	9892: gap of unknown length
*	9893	13224: contig of 3332 bp in length
*	13225	13242: gap of unknown length
*	13243	17573: contig of 4331 bp in length
*	17574	17591: gap of unknown length
*	17592	21248: contig of 3657 bp in length
*	21249	21266: gap of unknown length
*	21267	24477: contig of 3211 bp in length
*	24478	24495: gap of unknown length
*	24496	28851: contig of 4356 bp in length
*	28852	28869: gap of unknown length
*	28870	32392: contig of 3523 bp in length
*	32393	32410: gap of unknown length
*	32411	36463: contig of 4053 bp in length
*	36464	36481: gap of unknown length
*	36482	40605: contig of 4124 bp in length
*	40606	40623: gap of unknown length
*	40624	45997: contig of 5374 bp in length
*	45998	46015: gap of unknown length
*	46016	51508: contig of 5493 bp in length
*	51509	51526: gap of unknown length
*	51527	56900: contig of 5374 bp in length
*	56901	56918: gap of unknown length
*	56919	63213: contig of 6295 bp in length
*	63214	63231: gap of unknown length
*	63232	67639: contig of 4408 bp in length
*	67640	67657: gap of unknown length
*	67658	72522: contig of 4865 bp in length
*	72523	72540: gap of unknown length
*	72541	77773: contig of 5233 bp in length
*	77774	77791: gap of unknown length
*	77792	85251: contig of 7460 bp in length
*	85252	85269: gap of unknown length
*	85270	91567: contig of 6298 bp in length
*	91568	91585: gap of unknown length
*	91586	100917: contig of 9332 bp in length
*	100918	100935: gap of unknown length
*	100936	108346: contig of 7411 bp in length
*	108347	108364: gap of unknown length
*	108365	119883: contig of 11519 bp in length
*	119884	119901: gap of unknown length
*	119902	132327: contig of 12426 bp in length
*	132328	132345: gap of unknown length
*	132346	145829: contig of 13484 bp in length
*	145830	145847: gap of unknown length
*	145848	159854: contig of 14007 bp in length
*	159855	159872: gap of unknown length
*	159873	179103: contig of 19231 bp in length
*	179104	179121: gap of unknown length
*	179122	200644: contig of 21523 bp in length.

FEATURES Location/Qualifiers
source 1. .200644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-574H7"
BASE COUNT 48782 a 50967 c 50539 g 49852 t 5.4 others
ORIGIN

Query Match 10.8%; Score 383.4; DB 45; Length 200644;
Best Local Similarity 85.0%; Pred. No. 3.2e-66;
Matches 397; Conservative 0; Mismatches 66; Indels 4; Gaps 2;

```
Qy   2614 ggctgggccaggcacgatggctcacgcctataatcccagcactttgggaggccaa 2672
Db 131597 GGGGGTGGCGGGCACTGCGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCCAAAGCGGG 131656

Qy   2673 gcggatcacctgaggtcaggagtttgagaccagcctggccaacatggtgaaaccctgtctct 2732
Db 131657 GTAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTC 131716

Qy   2733 ctactaaaaatacaaaaattagccgggcgtggtggcgggcgcctgtaatcccagctactt 2792
Db 131717 CTACTAAAAATACAAAAATCAGCTAGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTC 131776

Qy   2793 gggaggctgaggcaggagaatcgcttgaacccgggaggcggaggttgcagtgagccgaga 2852
Db 131777 AGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGA 131836

Qy   2853 tcgtgccactgcactccagcctgggcgacagagcgagactccgtctcaaaaaaaaaaaaaa 2912
Db 131837 TCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAA 131896

Qy   2913 aataa---tcagggcatgagtggctctgcctgtaatcccagcactctgggaggctgaggct 2969
Db 131897 AAAAAAAGGCCGGGTGAGGTGGCTCACGCCTGTAATCACAGCACTTTGGGAGGCCGAGGCG 131956

Qy   2970 gggtggatcacctgaggtcaggagttcgagaccagcctggtgaaacatggtgaaaccccgt 3029
Db 131957 GGGCGGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTC 132016

Qy   3030 ctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
Db 132017 CTCTACTAAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTG 132063
```

```
RESULT  12
AC005972/c
LOCUS       AC005972   183451 bp    DNA             PRI       26-NOV-1998
DEFINITION  Homo sapiens chromosome 17, clone  hRPK.180_P_8, complete sequence.
ACCESSION   AC005972
VERSION     AC005972.1  GI:3928123
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183451)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.180_P_8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183451)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,P., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 183451)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,P., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 26, 1998 this sequence version replaced gi:3924658.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
FEATURES             Location/Qualifiers
     source          1. .183451
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone=" hRPK.180_P_8"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   complement(3. .53)
                     /rpt_family="AluJ/F(8)AM"
     repeat_region   64. .131
                     /rpt_family="Tigger2"
     repeat_region   complement(132. .441)
                     /rpt_family="AluSq"
     repeat_region   442. .664
                     /rpt_family="Tigger2"
     repeat_region   complement(665. .1216)
                     /rpt_family="L1MC2"
     repeat_region   complement(1217. .1526)
                     /rpt_family="AluSx"
     repeat_region   complement(1527. .1645)
                     /rpt_family="L1MC2"
     repeat_region   complement(1677. .1954)
                     /rpt_family="AluJo"
     repeat_region   complement(1966. .2046)
                     /rpt_family="L1MC2"
     repeat_region   complement(2064. .2156)
                     /rpt_family="Charlie1a"
     repeat_region   complement(2175. .2313)
                     /rpt_family="Charlie1a"
     repeat_region   complement(2547. .2592)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(2593. .2868)
                     /rpt_family="AluJb"
     repeat_region   2869. .2887
                     /rpt_family="AT_rich"
     repeat_region   complement(2973. .3221)
                     /rpt_family="AluJo"
     repeat_region   3366. .3689
                     /rpt_family="L1MEc"
     repeat_region   3690. .3850
                     /rpt_family="MLT1A2"
     repeat_region   3851. .4155
                     /rpt_family="AluSx"
     repeat_region   4156. .4365
                     /rpt_family="MLT1A2"
     repeat_region   4366. .4936
                     /rpt_family="L1MEc"
     repeat_region   complement(4937. .5237)
                     /rpt_family="AluSp"
     repeat_region   5238. .5624
                     /rpt_family="L1MEc"
     repeat_region   5631. .6182
                     /rpt_family="L1MEc"
     repeat_region   complement(6183. .6207)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(6211. .6482)
                     /rpt_family="AluSx"
     repeat_region   6483. .7290
                     /rpt_family="L1MEc"
     repeat_region   7288. .8016
                     /rpt_family="L1MD1"
     repeat_region   8017. .8302
                     /rpt_family="AluJb"
     repeat_region   8303. .8751
                     /rpt_family="L1MD1"
     repeat_region   8816. .9455
                     /rpt_family="L1MD1"
     repeat_region   10240. .10267
                     /rpt_family="AT_rich"
     repeat_region   10386. .10920
                     /rpt_family="L1MD1"
     repeat_region   10923. .11718
                     /rpt_family="L1PA14"
     repeat_region   11719. .12013
                     /rpt_family="AluYb8"
     repeat_region   12014. .12498
                     /rpt_family="L1PA14"
     repeat_region   12499. .12904
                     /rpt_family="L1MD1"
     repeat_region   12905. .13217
                     /rpt_family="AluSq"
     repeat_region   13228. .13539
                     /rpt_family="AluSx"
     repeat_region   13540. .13992
                     /rpt_family="L1MD1"
     repeat_region   14076. .14377
                     /rpt_family="AluY"
     repeat_region   complement(14708. .15273)
```

```
                     /rpt_family="L1MC4"
     repeat_region   complement(15429. .15604)
                     /rpt_family="AluSg/x"
     unsure          complement(15816. .15980)
                     /note="Single-stranded coverage."
     repeat_region   16294. .16603
                     /rpt_family="AluSx"
     repeat_region   complement(16707. .17214)
                     /rpt_family="L2"
     repeat_region   complement(17215. .17381)
                     /rpt_family="AluY"
     repeat_region   complement(17382. .17713)
                     /rpt_family="AluY"
     repeat_region   complement(17714. .17852)
                     /rpt_family="AluY"
     repeat_region   complement(17856. .18154)
                     /rpt_family="AluSg"
     repeat_region   complement(18155. .18278)
                     /rpt_family="L2"
     repeat_region   complement(18483. .18779)
                     /rpt_family="AluSg"
     repeat_region   18857. .19114
                     /rpt_family="AluSc"
     unsure          complement(19096. .19126)
                     /note="Single-stranded coverage."
     unsure          19547. .19632
                     /note="Single-stranded terminator coverage."
     repeat_region   20375   20678
                     /rpt_family="AluSx"
     repeat_region   21441. .21563
                     /rpt_family="L2"
     repeat_region   21650. .21853
                     /rpt_family="AluSg/x"
     repeat_region   complement(21859. .22046)
                     /rpt_family="THE1B"
     repeat_region   22297   22622
                     /rpt_family="MSTA"
     repeat_region   23381. .23596
                     /rpt_family="AluJ/FLAM"
     repeat_region   24249. .24558
                     /rpt_family="AluSx"
     repeat_region   complement(25629   25704)
                     /rpt_family="MIR"
     repeat_region   complement(25762. .26051)
                     /rpt_family="AluSx"
     repeat_region   26574   26853
                     /rpt_family="AluSc"
     repeat_region   26854. .26894
                     /rpt_family="(TAAA)n"
     repeat_region   27131. .27435
                     /rpt_family="AluSq"
     repeat_region   complement(27437. .27642)
                     /rpt_family="MER2"
     repeat_region   complement(27832. .27896)
                     /rpt_family="MER58"
     repeat_region   complement(28231. .28270)

  Query Match            10.8%;  Score 302;  DB 40,  Length 183451;
  Best Local Similarity   84.4%;  Pred. No. 6.7e-66;
  Matches  410;  Conservative    0;  Mismatches   70;  Indels    6;  Gaps    3;

Qy   2592 tctctggaagtatttaagaaaaggctgggccaggcacgatggctcac-gcctgtaatccc 2650
          ||  ||     || ||| |||||  |||| |||||  |||||||| ||||||||||
Db 100635 TGTGIGGGGCAIITAAASIAAAGGGCASSYCIGGCACACTGGCTCACACCCTGTAATCCC 100576

Qy   2651 agcactttgggaggccgaggcaggcggatcacctgaggtgaggagtttgagaacagcctg 2709
          || ||||||||||| |||||||||  |||| ||| || || ||||  ||| ||||||||
Db 100575 AGCACTTTGGGAGGCTGAGGCAGGCAAATCACTTGAAGTCAGAAGTTCAAGACCAGCCTG 100516

Qy   2710 gccaacatggtgaaaccctctctctactaaaaatacaaaaattagccaggcgtggtggca 2769
          |||||| |||||||| | |||||||||| |||||||||||||||  ||| ||||| ||
Db 100515 GGCAACAGGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGGCATGGTGACA 100456

Qy   2770 ggtgcctgtaatcccagctacttgggaggctgaggcatgagaatcgcttgaac-tgggag 2828
          ||  |||||||||||||||| | | | | | | | | | | | | | | | | | | | |
Db 100455 GGCAGCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTAAACCTGGGAG 100396

Qy   2830 gcagaggttacagtgagccaagattgtgccactgcactccagcctgggcgacagagcgaga 2889
          || ||||| ||||||| || | | | | | | | | | | | | | | | | | | | |
Db 100395 GCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGCGAG 100336

Qy   2890 actctgtctcaaaaaaaaataaaaaaatcaatcagggcatcagtcccctatgcctgtagtcc 2949
          ||| ||||||||||    || | | | | | | | | | | | | | | | | | | | |
Db 100335 ACTCAGTCTCAAAAA----AAIAAAAGCCCCAGCCAGCCTCCCCTCAGCCCTCTAATCCCN 100280

Qy   2950 gcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcgagaccagcctgac 3009
          ||||| |||||||||||||||| | | | | | | | | | | | | | | | | | | |
Db 100279 GCACTTTGGGAGGCTGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGAC 100220

Qy   3010 tgaacatggcgaaaccccgtctctaaaaaatacaaaaattagccgggcatggtggtgc 3069
          |||||| | || |||| | | | | | | | | | | | | | | | | | | | |
Db 100219 CCAACATGGGGAAATCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCA 100160

Qy   3070 atgcct 3075
          ||||||
Db 100159 GTGCCT 100154


RESULT  13
AC007615
LOCUS       AC007615   185997 bp    DNA             HTG       20-MAY-1999
DEFINITION  Homo sapiens chromosome 16 clone 323K11, *** SEQUENCING IN PROGRESS
            ***, 44 unordered pieces.
ACCESSION   AC007615
VERSION     AC007615.1  GI:4878051
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185997)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Mark,O., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussell,M., Sutherland,R.,
            Han,C. and Deaven,L.
  TITLE     Sequencing of Human Chromosome 16q12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185997)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 185997)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Mark,O., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussell,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 44 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      517: contig of 517 bp in length
            *                    gap of unknown length
            *      518     1027: contig of 510 bp in length
            *                    gap of unknown length
            *     1028     1733: contig of 706 bp in length
```

```
*                    gap of unknown length
*       1734    2167: contig of 434 bp in length
*                    gap of unknown length
*       2168    2695: contig of 528 bp in length
*                    gap of unknown length
*       2696    3519: contig of 824 bp in length
*                    gap of unknown length
*       3520    4811: contig of 1292 bp in length
*                    gap of unknown length
*       4812    5834: contig of 1023 bp in length
*                    gap of unknown length
*       5835    6922: contig of 1088 bp in length
*                    gap of unknown length
*       6923    7048: contig of 126 bp in length
*                    gap of unknown length
*       7049    8115: contig of 1067 bp in length
*                    gap of unknown length
*       8116    9016: contig of 901 bp in length
*                    gap of unknown length
*       9017   10056: contig of 1040 bp in length
*                    gap of unknown length
*      10057   10714: contig of 658 bp in length
*                    gap of unknown length
*      10715   11357: contig of 643 bp in length
*                    gap of unknown length
*      11358   12588: contig of 1231 bp in length
*                    gap of unknown length
*      12589   13474: contig of 886 bp in length
*                    gap of unknown length
*      13475   15190: contig of 1716 bp in length
*                    gap of unknown length
*      15191   16581: contig of 1391 bp in length
*                    gap of unknown length
*      16582   18148: contig of 1567 bp in length
*                    gap of unknown length
*      18149   19342: contig of 1194 bp in length
*                    gap of unknown length
*      19343   21196: contig of 1854 bp in length
*                    gap of unknown length
*      21197   22929: contig of 1732 bp in length
*                    gap of unknown length
*      22929   24423: contig of 1495 bp in length
*                    gap of unknown length
*      24424   27038: contig of 2615 bp in length
*                    gap of unknown length
*      27039   29016: contig of 1978 bp in length
*                    gap of unknown length
*      29017   32592: contig of 3576 bp in length
*                    gap of unknown length
*      32593   34056: contig of 1464 bp in length
*                    gap of unknown length
*      34057   37214: contig of 3158 bp in length
*                    gap of unknown length
*      37215   40156: contig of 2942 bp in length
*                    gap of unknown length
*      40157   43893: contig of 3737 bp in length
*                    gap of unknown length
*      43894   46580: contig of 2687 bp in length
*                    gap of unknown length
*      46581   49899: contig of 3319 bp in length
*                    gap of unknown length
*      49900   55101: contig of 5202 bp in length
*                    gap of unknown length
*      55102   60409: contig of 5308 bp in length
*                    gap of unknown length
*      60410   70936: contig of 10527 bp in length
*                    gap of unknown length
*      70947   81300: contig of 10354 bp in length
*                    gap of unknown length
*      81301   89022: contig of 7722 bp in length
*                    gap of unknown length
*      89023   99895: contig of 9873 bp in length
*                    gap of unknown length
*      98896  109265: contig of 10370 bp in length
*                    gap of unknown length
*     109266  118381: contig of 9116 bp in length
*                    gap of unknown length
*     118382  132765: contig of 14384 bp in length
*                    gap of unknown length
*     132766  149210: contig of 16445 bp in length
*                    gap of unknown length
*     149211  185997: contig of 36787 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="528K16"
                     /chromosome="16"
BASE COUNT    47380 a  46268 c  45654 g  46667 t     28 others
ORIGIN


  Query Match           10.8%;  Score 331,  DB 33,  Length 185997;
  Best Local Similarity   82.0%;  Pred. No. 1.1e-65;
  Matches  393;  Conservative    0;  Mismatches   85;  Indels    1;  Gaps    1;

Qy    2599 aagtatttaagaaaaggctgggccaggcacgatggctcacgcctgtaatcccagcactttg 2658
           |||| |  || | ||     ||||  |||  | ||||||| ||||||||||||| | |||
Db   33137 AAGTCTAAAAAATAAATGGGGGTGGGGTGGTGGCGCATGCCTGTAATCCCAGCACTTTG 33196

Qy    2659 gggaggccgaggcaggcggatcacctgaggtcaggagtttgagaccagcctgaccaacat 2717
           |||||||| |||  ||| |||||||||| || |||||||  ||| ||||||| ||||||
Db   33197 GGGAGGCCAAGGTGGGCAGATCACCTGAAGTCAGGAGTTCAAGACCAGCCTGGCCAACAT 33256

Qy    2718 ggtgaaacctcatctctactaaaaatacaaaaattagccaggcgtggtggcgcgcctgtag 2777
           |||||||     |||||||||||||||||||||||||||||  ||| |||||  |||| |
Db   33257 AGTGAAACTCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCGGGTGCCTG 33316

Qy    2778 taatcccagctacttgggaggctgaggcaggagaatcacttgaacctgggaggcggaggt 2837
           |||||||||||||||  |||||||||||||| ||||||  ||| |||| | | | ||||
Db   33317 TAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGACGGAGGT 33376

Qy    2838 tacagtgagccaagatcgtgccactgcattccagcctgggcgacagagcgagactccgtct 2897
           | ||||||||||||||  |||||||| |||||| |||| |||||||||| | |||||||
Db   33377 TGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCCGTC 33436

Qy    2898 tcaaaaaaaaataaaaaataatcagggcacattggctcatgcctgtaatcccagcactttg 2957
           ||||||| ||  ||| ||| | | | ||||||||||||| | | |||
Db   33437 TCAAAAATAAATAAACATAGGCCAGGTGCAGTGGCTCATGCCTATAATCCCAGCACTTTG 33496

Qy    2958 ggaggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctgaccaacatg 3017
           |||||||||  |   |||| | |||||| |||| ||||||||| | | | | | | ||||
Db   33497 AGAGGCTGAGGCAGAAAAATCACTTGAGGTCAGAAGTTCGAGACCAGCCTGGCCAACATG 33556

Qy    3018 gcgaaaccccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctgta 3076
           | ||| || | |||||| |||||| |||||||||| ||| |||||||||| ||||
Db   33557 GTGAACCCTCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCGCATGCCTG 33615


RESULT  14
HUAC002544
LOCUS       HUAC002544 156641 bp    DNA             PRI       22-NOV-1999
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-761H5, complete
            sequence.
ACCESSION   AC002544
VERSION     AC002544.1  GI:3337382
KEYWORDS    HTG
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156641)
  AUTHORS   Loftus,B.J., Kim,U.J., Sheddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
```

```
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
REFERENCE   2  (bases 1 to 156641)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-761H5
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 156641)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            mdadams@tigr.org
REFERENCE   4  (bases 1 to 156641)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:2959563.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.  The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://genomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1..156641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p11.2"
                     /clone="A-761H5"
     STS             7426..7531
                     /note="44231, SHGC-36058, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G30059"
     gene            27016..108426
                     /gene="A-761H5.3"
                     /pseudo
     STS             27283..27567
                     /gene="A-761H5.3"
                     /note="25260, CHLC.GGT16G06, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G16020"
     mRNA            complement(join(32539..32821,32917..33079,33182..33269,
                     33358..33423,33522..33681,35998..36361,40957..41181,
                     41407..41597,43190..43283,43419..43544,43739..43880,
                     44649..44752,44877..45036,52772..52899,53278..53387,
                     53516..53570,53760..53885,54108..54297,54440..54505,
                     56420..56550,56689..56743))
                     /gene="A-761H5.1"
     gene            complement(32539..56743)
                     /gene="A-761H5.1"
     CDS             complement(join(32739..32821,32917..33079,33182..33269,
                     33358..33423,33522..33681,35998..36361,40957..41181,
                     41407..41597,43190..43283,43419..43544,43739..43880,
                     44649..44752,44877..45036,52772..52899,53278..53387,
                     53516..53570,53760..53885,54108..54297,54440..54505,
                     56420..56550))
                     /gene="A-761H5.1"
                     /codon_start=1
                     /product="Translation initiation factor eIF-p110"
                     /protein_id="AAC27426.1"
                     /db_xref="GI:3337393"
                     /translation="MSPEFIGSDSESESSLSGEELVTKPVGGNYGKQPLLLSEDEED
                     IKPVVPSAKDKPPEELTNLIPGIPNAMKIPDVTKCLEEPELLGKAYGKAKSIVDKEGV
STS                  57091..57208
                     /gene="A-761H5.3"
                     /note="77608, STS:USA-2.gi:2484[illegible], Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
mRNA                 join(<66573..66995,67730..68018)
                     /gene="A-761H5.2"
gene                 <66573..>68018
                     /gene="A-761H5.2"
CDS                  join(66573..66995,67730..68018)
                     /gene="A-761H5.2"
                     /codon_start=1
                     /product="CDC37-like gene"
                     /protein_id="AAC27427.1"
                     /db_xref="GI:3337394"
                     /translation="MAIGRSTTSLRITVISVSWRAEAKCWIIAGAAILSDEEERHPN
                     THPANLFPWQPLARVELMEIPKPEVEPIPAPIVNRVPMALLSAPLRELEVSEGARVE
                     VEWLHEELSWEQKLEEMPRPPMTDNVFTLPMSDPSLIEMEIPAAPLMLQDIDAGLR
                     VPNSEVSEAKEGEEECSGIGCWKLFPRQASMSVRCPPQLLLPPVCKPLSAPFQKTKID
                     IISPAPGPLHPRCSP"
mRNA                 complement(join(<109422..110066,111426..111461,
                     111571..111631,115712..115967,116020..116119,
                     123403..>123468))
                     /gene="A-761H5.4"
gene                 complement(<109422..>123468)
                     /gene="A-761H5.4"
CDS                  complement(join(109422..110066,111426..111461,
                     111571..111631,115712..115967,116020..116119,
                     123403..123468))
                     /gene="A-761H5.4"
                     /codon_start=1
                     /product="Unknown gene product spliced form-1"
                     /protein_id="AAC27428.1"
                     /db_xref="GI:3337395"
                     /translation="MRIRFWVIWLLLCFISHQTTVSPLKTIPWSWNSHDGSMDVQQ
                     RAWFSNKSPQFSLRSIDWRIRKRVSSAGNRILKALVILLKRWLDIRLRAWKIKKKKVT
                     TKINRHDKINGKRPTAPKLMVTLRKQDLFFRAEDVRKIFISAKRILDNWVRMAAAE
                     HRHSSGLFYWLLLIALILRNFMSR,FIIILVDDILNSLSLKIPPEDLLIPLPPSVD
                     DNIKEIPLAPLPPSPLPPSVDLNLREDLFVPLPPSPLPLSVDLNLKEDLFVPLPPSPL
                     FPSVDDNLKTFFLATQEAPVFKPPPPPKWRVISVELSPKWKRRRVVEVELSPKWKRQR
                     EAEAQQLFKPKRRRLSKLKIRHDILAWAIKINP"
STS                  109425..109587
                     /gene="A-761H5.4"
                     /note="40991, SHGC-32146, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G27136"
mRNA                 complement(join(110411..111461,111571..111631,
                     115712..115967,116020..116119,123403..>123468))
                     /gene="A-761H5.5"
gene                 complement(110411..>123468)
                     /gene="A-761H5.5"
CDS                  complement(join(111219..111461,111571..111631,
                     115712..115967,116020..116119,123403..123468))
                     /gene="A-761H5.5"
                     /codon_start=1
                     /product="Unknown gene product spliced form-2"
                     /protein_id="AAC27429.1"
                     /db_xref="GI:3337396"
                     /translation="MRIRFWVIWLLLCFISHQTTVSPLKTIPWSWNSHDGSMDVQQ
                     RAWFSNKSPQFSLRSIDWRIRKRVSSAGNRILKALVILLKRWLDIRLRAWKIKKKKVT
```

```
                     IKINPHDKINGKPKTARKQKVSGPAQELRPPAEDYHKCKIPPSARKPLCNWVSLFVFL
                     APEHSLPGQDMDTFFSLQICALALQREMAEPKAAYRHHSPIPVGNPVVQKHLHPHPVG
                     PLI"
     STS             121375. .121480
                     /gene="A-761H5.5"
                     /note="44231, SHGC-36059, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS.[illegible]"
     mRNA            complement(join(<130544   130663,130765. .130905,
                     135131. .135224,135353. .135408,135503. .135571,
                     135652. .135698,137020. .137132,139362. .139505,
                     139593. .139665,140471. .140556,140677. .140756,
                     141606. .141677,142305. .142401,144492. .144570,
                     144724. .144946))
                     /gene="A-761H5.6"
     gene            complement(<130544. .144946)
                     /gene="A-761H5.6"
     CDS             complement(join(130544. .130663,130765. .130905,
                     135131. .135224,135353. .135408,135503. .135571,
                     135652. .135698,137020. .137132,139362. .139505,
                     139593. .139665,140471. .140556,140677. .140756,
                     141606. .141677,142305. .142401,144492. .144570,
                     144724. .144769))
                     /gene="A-761H5.6"
                     /codon_start=1
                     /product="CLN"
                     /protein_id="AAC27430.1"
                     /db_xref="GI:3337387"
                     /translation="MGGCAGSRPRPSDSEGEETVPEPPLPLLDHQGAHWKNAVGFWLL
                     GLCNNFSYVVMLSAAHDCLSHRTSGNQSHVDPGPIPIPHNSSSPFDCNSVSTAAVLL
                     ADILPTLVIKLLAPLGLHLLPYSPRVLVSGICAAGSFVLVAFSHSVGTSLCGVVFASI
                     SSGLGEVTFLSLTAFYPPAVISWWSSGTGGAGLLGALSYLGLTQAGLSPQQTLLSMLG
                     IPALLLASYFLLLTSPEAQDPRGEEEAESAARQPLIRTEAPESKPGSSSSLSLPERWT
                     VFKGILWYIVPLVVVYFAFYPINQGLFELLFFWNTSLSHAQQYPWYQMLYQAGVFASP
                     SSLRCCRIRFTWALALLQCLNLVFLLADVWFGFLPSIYLVFLIILYEGLLGGAAYVNT
                     FHNIALETSDEHREFAMAAIQISDTLGISLSGLLALPLHDFLCQLS"
BASE COUNT    41602 a  38831 c  38211 g  37878 t     59 others
ORIGIN


  Query Match           10.8%;  Score 331;  DB 40;  Length 156641;
  Best Local Similarity   82.0%;  Pred. No. 1.1e-65;
  Matches  393;  Conservative    0;  Mismatches   85;  Indels    1;  Gaps    1;

Qy    2599 aagtatttaagaaaaggctgggccaggcacgatggctcacgcctgtaatcccag-acttt 2657
           |||| | || | ||    |||| ||| |  | |||||| |||||||||||||| ||| | |||||
Db   93582 AAGTCTAAAAAAATAAATGGGGGCTGGGCGTGGTGGCTCATGCCTGTAATCCCAGCACTCT 93641

Qy    2658 gggaggccgaggcaggcggatcacctgaggtcaggagtttgagaacagcctggccaacat 2717
           ||||||| ||| ||| |||||||||| || ||||||  ||| ||||||||||||||||||
Db   93642 GGGAGGCCAAGGTGGGCAGATCACCTGAAGTCAGGAGTTCAAGACCAGCCTGGCCAACAT 93701

Qy    2718 gatgaaaccccatctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctg 2777
           |||||||  |||||||||||||||| || |||||  ||| ||||||| ||||||||||||
Db   93702 AGTGAAACCTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCGGGTGCCTG 93761

Qy    2778 taatcccagctacttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggt 2837
           |||| | |||||  ||||||||||| | ||| ||| |||| | |||| | |||||||||
Db   93762 TAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGACGGAGGT 93821

Qy    2838 tacagtgagccaagatcgtgccactgcattccagcctgggcgacagagcaagactctgtc 2897
           | |||||||||||||  |||||||| |||||  |||| |||||||| ||||||||||||
Db   93822 TCCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCTGTC 93881

Qy    2898 tcaaaaaaaaataaaaaataatcagggcagagtggctcatgcctgtaatcccagcactctg 2957
           | | | | ||| ||  ||  || |||||||| |||||||| ||| | |||||||||||||
Db   93882 TCAAAAAIAAAIAAACAIAGGCCAGGTGCAGTGGCTCATGCCTATAATCCCAGTACTTTG 93941

Qy    2958 ggaggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatg 3017
           |||||||||| |  ||||| ||||||||  ||||| ||||||||||||  ||||||||||
Db   93942 AGAGGTTGAGGCAGAAAAATCATTTGAGCCCAGAAGTTCGAGACCAGCCTGGCCAACATG 94001

Qy    3018 gcgaaaccccgtctctactaaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           | | | |  |||||| |||| |||| ||  |||| ||| ||||||||||||||||||
Db   94002 GTGAACCCTCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTG 94060


RESULT  15
HSJ189G13
LOCUS       HSJ189G13  129294 bp    DNA             HTG       03-DEC-1999
DEFINITION  Homo sapiens chromosome 20 clone RP1-189G13, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL121916
VERSION     AL121916.6  GI:6523724
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129294)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 3, 1999 this sequence version replaced gi:6469213.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.  Order of segments is not known, 800 n's separate
            segments. Contig_ID: 00483  Length: 56195bp
            Contig_ID: 00651  Length: 14042bp
            Contig_ID: 00773  Length: 38840bp
            Contig_ID: 01169  Length: 4797bp
            Contig_ID: 01486  Length: 1395bp
            Contig_ID: 01579  Length: 1292bp
            Contig_ID: 01680  Length: 6097bp
            Contig_ID: 01894  Length: 1036bp.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .129294
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP1-189G13"
                     /clone_lib="RPCI-1"
BASE COUNT    36075 a  25794 c  25357 g  36466 t    5602 others
ORIGIN


  Query Match           10.8%;  Score 330.8;  DB 33;  Length 129294;
  Best Local Similarity   85.1%;  Pred. No. 1.3e-65;
  Matches  393;  Conservative    0;  Mismatches   67;  Indels    2;  Gaps    2;

Qy    2617 tgggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcgggcgg 2675
           | ||||||| ||  ||||||| | |||||||||||| |||||||||| | | | | | |
Db   80586 TAGGCCAGGTACAGTGGCTCATGACTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCGG 80645

Qy    2676 gatcacctgaggtgaggagtttgagaccagcctggccaacatggtgaaaccctgtctctg 2735
           |||||| |||||| ||||||| |||| ||||||||||||||||| ||||||| | ||||||
Db   80646 GATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGATGAAACCCCGTCTCTA 80705

Qy    2736 ctaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctactgggg 2795
           |||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||
Db   80706 CTAAAAATACAAAAATTAGCCAGGCGTGGTGGCACACGCCTGTAATCCCAGCTACTGGGG 80765

Qy    2796 aggctgaggcatgagaatcacttaaacctgaggaggcagaggttgcagtgagccgagatc 2855
           |||||||||| |||||||||| ||| || |||||||| | | || || ||||||
Db   80766 AGGCTGAGGCGGGAGAATCACTTGAACATGGGAGGCAGAGGTTGCATTGAGCTGAGATC 80825

Qy    2856 tgccactgcattccagcctgggcgacagagcgagactctgtctcaaaaaaaaaaaaaaaa 2915
           |||||||||| |||  |||||||||||| | |||| | | |||||||||||||||||||
Db   80826 TGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAA 80885
```

```
Db  80826 NGGCACTGGACTCCCACCTGGGCGACAGAGGGAAACTCAGTCTCAAAAAAAAAAAAAAAA 80885

Qy   2916 aa-tcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggtgggtg 2974
          ||  :  ||||  ||||||||||||||||||| || ||  | ||||||||||||||| ||
Db  80886 AAGGCCAGGCATGGTGGCTCATGCCTGTAATCCCGGCGATTTGGGAGGCTGAGGTGGTTG 80945

Qy   2975 gatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccccgtctcta 3034
          |||||  ||||||||| |||||  |||||||||| ||  |||||| ||||||| ||||||
Db  80946 GATCATTTGAGGTCACGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTA 81005

Qy   3035 ataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
          |||||||  |||||||| ||||||||.|.|||  |||
Db  81006 CTAAAAATATGAAAATTAGCTGGGCATGGTGGTGCACACCTG 81047


Search completed: February  9, 2000, 07:04:08
Job time: 9562 sec
```

```
OM of: US-09-174-937-2 to: GenEmbl *   -o=_format   pfs

Date: Feb 9, 2000  4:43 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -S=/cgn2_1/USPTO_spool/US09174937/runat__1220[illegible]/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -QGAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=1.000 -XGAPOP=10.000 -XGAPEXT=0.100 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-174-937-2
Query length: 371
Database: GenEmbl.*
Database sequences: 821193
Database length: 1518192014
Search time (sec): 1185.490000


score_list:
Sequence              Strd Orig    %Score   EScore Len   ! Documentation ..
gb_pri:AB002365         +  992.50 1713.20  2.1e-87   5654  ! AB002365 Human mRNA for KIAA0
gb_ro:AF035207          +  949.50 1645.26  1.2e-83   2213  ! AF035207 Mus musculus Nip21 mR
gb_pat:AR028453         +  935.50 1621.13  1.3e-82   1386  ! AR028453 Sequence 22 from patu
gb_pr2:HSU15173         +  935.50 1621.62  2.7e-82   2382  ! U15173 Homo sapiens BCL2/adeno
gb_htg2:AC011488        -  482.00  779.02  2.3e-35  179301  ! AC011488 Homo sapiens chromo
gb_htg5:AC015296        -  436.00  706.46  2.5e-31   79170  ! AC015296 Drosophila melanogas
gb_pri:HSRHOGAPX        -  295.50  482.14  7.8e-19   1430  ! Z23024 H.sapiens rhoGAP protei
gb_pr1:HSU02570         -  295.50  472.49  2.4e-18   3845  ! U02570 Human CD042 GTPase-acti
gb_htg6:AC009556        -  279.00  422.08  1.7e-15  169798  ! AC009556 Homo sapiens chromo
gb_htg6:AC009556        +  254.00  378.06  4.9e-13  169798  ! AC009556 Homo sapiens chromo
gb_htg5:AC014963        -  235.00  349.39  1.9e-11  106121  ! AC014963 Drosophila melanoga
gb_pr2:HSU62794         -  172.00  284.40  9.1e-08   1378  ! U62794 Human CD042 GAP-related
gb_htg1:CBY6285         +  170.00  223.99  0.0002   310565  ! AL023633 Caenorhabditis eleg
gb_in1:CEW02B12         +  170.00  245.40  1.1e-05   34205  ! Z66521 Caenorhabditis elegans
gb_in2:AF091997         +  142.50  227.53  0.0001    2234  ! AF091997 Plasmodium falciparum
gb_in2:AC004283         +  143.00  196.23  0.0143   80095  ! AC004283 Drosophila melanogas
gb_ro:MMY17851          +  138.00  207.48  0.0016    2604  ! Y17851 Mus musculus mRNA for g
gb_in2:AF091996         +  130.00  208.06  0.0014    1464  ! AF091996 Plasmodium falciparum
gb_htg3:AC011488        -  129.00  157.37  0.9636  179301  ! AC011488 Homo sapiens chromo
gb_pat:I70400           -  114.00  169.45  0.0047    4080  ! I70400 Sequence 1 from patent
gb_pat:I70401           -  114.00  175.31  0.0966    2295  ! I70401 Sequence 3 from patent
gb_pl1:SCU12137         -  114.00  167.35  0.2679    5015  ! U12137 Saccharomyces cerevisia
gb_pl1:SCU14974         -  114.00  169.45  0.2047    4080  ! U14974 Saccharomyces cerevisia
gb_pl1:SCU28158         -  114.00  167.62  0.2587    4883  ! U28158 Saccharomyces cerevisia
gb_pl2:YSCU10556        -  114.00  146.43  3.92     33203  ! U10556 Saccharomyces cerevisi
gb_pl2:SPAC6F6          -  105.50  130.26  31.18    44092  ! Z98981 S.pombe chromosome I c
gb_pl2:ATAC005727       -  103.00  117.71  [illegible]  107037  ! AC005727 Arabidopsis thalian
gb_pl1:SC8261X          -  102.50  129.53  31.27    28201  ! Z49809 S.cerevisiae chromosom
gb_pat:I32737           -  101.50  152.41  1.92      2704  ! I32737 Sequence 15 from patent
gb_pat:I39712           -  101.50  152.41  1.92      2704  ! I39712 Sequence 15 from patent
gb_pl1:SCCHXIV43        +  101.50  123.36  75.57    43481  ! Z46843 S.cerevisiae chromosom
gb_pl1:SCCXIV39K        -  101.50  124.79  60.95    37800  ! Z69382 S.cerevisiae 37,800 bp
gb_pl1:SCYNL120W        -  101.50  142.27  5.81      6586  ! Z71402 S.cerevisiae chromosome
gb_pl1:SCYNL127W        -  101.50  148.82  2.89      3563  ! Z71403 S.cerevisiae chromosome
gb_pl2:SPBC29A10        -  101.50  123.56  73.61    42612  ! AL034463 S.pombe chromosome I
gb_pl2:SPU38522         -  101.50  150.42  2.85      3049  ! U38522 Schizosaccharomyces pom
gb_pl2:SPU40378         +  101.50  153.06  1.68      2349  ! U40378 Schizosaccharomyces pom
gb_htg2:AC007805        -  101.50  112.92  299.37  101315  ! AC007805 Drosophila melanoga
gb_htg7:AC018284        -  101.50  119.90  199.89   57392  ! AC018284 Drosophila melanogas
gb_htg3:AC007301        -  100.00  107.29  522.99  162666  ! AC007301 Drosophila melanoga
gb_pl1:AP016876         +   99.50  119.60  122.37   44499  ! AB016876 Arabidopsis thaliana
gb_pl1:S54965           -   99.50  153.88  1.51      1539  ! S54965 pgk=3-phosphoglycerate

gb_in1:CELF55A6         -   99.50  [illegible]  91.12  [illegible]  ! [illegible] Caenorhabditis el
gb_in1:PALCOL1A         -   99.00  149.35  2.69      2197  ! M26242 P.lividus alpha co
gb_pr1:AB002804         -   98.50  138.56  10.76     5811  ! AB002804 Homo sapiens mRN

seq_name: gb_pr1:AB002365

seq_documentation_block:
LOCUS       AB002365     5654 bp    mRNA            PRI       19-FEB-1999
DEFINITION  Human mRNA for KIAA0367 gene, partial cds.
ACCESSION   AB002365
VERSION     AB002365.1  GI:2224674
KEYWORDS    KIAA0367.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone:pBluescriptII SK
            plus clone:HH0141.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5654)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Chiba,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1, 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:ndnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3931,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Chiba,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1..5654
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH0141"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..2151
                     /gene="KIAA0367"
     CDS             <1..2151
                     /gene="KIAA0367"
                     /codon_start=1
                     /protein_id="BAA21212.1"
                     /db_xref="GI:2224675"
                     /translation="ALPLSDPHLSIENFALVPLALLASLPTLLDISEAAPDHSFSDASG
                     LNTSTGTIDDMSPLTLSEDHTTPVTSLSKPLPSSSLASMSLPELVMSQNIDEDLL
                     REPRHFIYGSFPRELLSLKLSIAVTPAPTLNYLTEPAQSAETTEPASSPEDESLGC
                     RAAEIVLSALPGKSPQNLADTKNFLPQSLLAVLNHIPELPEQVVLPVGASNILSPSN
                     VDWEVETDNSLLPASDIDFSNGAGKLLLLLLDKIPKLQLPKSLLKEERCIEKN
                     EDRHALHMDYIGVNPLLNSHGAFEIILDLSIAELLLTVSSAPISLGLILASFPDTC
                     QPASLNERKSLCALPNPATSLTPSSFLCLALLLWNHLNCLCLPTSLLKALSGPGWS
                     GKIVEPFSELGLSESPLLLLLEMAFLSSLALLLASGFVSPLASKSPSRAGPDAVTH
                     DSEWEMLSFQPVCKNMIGCTDMEEETFLPGCTPTSTPNGLLSFFVGMCIPPEEGVLS
                     PSAADMRFEFPNSLLDNLTHTRIKLLNKNLSLWLDSSLSLLNLSPDEIDINV
                     CELTPLEADLPLTGMLTNMKDSDLEGLLDPTMMSTFSMPLVVIGEQEQR
                     IDMRVIDFVRSVISPDTVSDLNKIVPVALPLSSAVLLLSSLNLFLYVISTLE
                     LMVAEDPMIVTLNSAIKRELAAASLLIGMLPIS"
BASE COUNT      1664 a    1288 c    1198 g    1504 t
ORIGIN


alignment_scores:
            Quality:  992.50            Length:    371
              Ratio:  3.422               Gaps:      11
 Percent Similarity:  97.907   Percent Identity:  95.648

alignment_block:
 US-09-174-937-2 x AB002365   ..
```

```
Align seq 1/1  to: AB002365  from: 1  to: 5654

     12 AsnValAspValLysGluGluTrp   Glu AspGluAspLeuProArgP 27
        ||   |:::|||  :::||       :::|||||:::
   1446 AACCAGGATATCAAGACCAAATGGACTACTGTCAGAGGATGTAGGAATGG 1495

     27 roLeuProGluGluThrGlyValGluLeuLeuGlySerProValGluAsp 43
        ..:::  |||  ||||||  |||        |||  .....:.
   1496 ACATCCCTTTGAAGAGGCGTG  CTGAGTCCAGTGCTGCAGACATG 1542

     44 ThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLysAr 60
           :::|||:|||||:::|||:::::||   :::|||  :::||
   1543 AGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAG 1592

     60 gLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerGluG 77
        |    |||:::|||||||:::|||||:::|||||||||||||||||||
   1593 AATCAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAG 1642

     77 lySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuAspIle 93
        |||:::|||||||||   |||||||::  ||||:::::|||||
   1643 GATCTATTCTCTCTGATGATAACTTGGACAGTCTAGATGAAATTGACATC 1692

     94 AsnValAspAspIleGluThrProAspGluThrAspSerLeuGluPheLe 110
        ||||||||||:::::::::|||||||||||||:::|||||:::|||:::
   1693 AATGTGGATGAACTTGATACCCCCGATGAAGCAGATTCTTTTGAGTACAC 1742

    110 uGlyAspGlyAspGluLeuGluTrpGluAspAspThrProValAlaThrA 127
          | :::                                   ||||
   1743 TGGCCATGAT                              . . . CCCACAG 1759

    127 laLysAsnMetProGlyAspSerAlaAspLeuPheGlyAspGlyThrThr 143
        ||:::::    |||:::::::...              :   |||:
   1760 CCAACAAAGATTCTGGCCAAGAGTCAGAGTCTACTCCAGAATATACGGCC 1809

    144 GluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrValIleIleGl 160
        |||:::            |||   ||||| ||| ||||||||:::|||||
   1810 GAAGATGGAACGGGAGGACAAC   CGGCTTTCTAGGACAGTGGTCATTGG 1856

    160 yGluGlnGluHisArgIleAspLeuHisMetIleArgProTyrMetLysV 177
        ||||||||||:::||||||||:..  ...:   ...||||||    ...:
   1857 AGAACAAGAGCAGCGCATTGACATGAAGGTCATCGAGCCCTACAGGAGAG 1906

    177 alValThrHisGlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleVal 193
        ||:::::::||||||||||||||||:::  ||   |||||||||||||||
   1907 TCATTTCTCACGGAGGATACTATGGGGACGGTCTAAATGCCATCATTGTG 1956

    194 PheAlaAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrIl 210
        ||||||||||||||||||||||||||||||       ||||||||||||::
   1957 TTTGCCGCCTGTTTTCTGCCAGACAGCAGTCGGGGGGATTACCACTATGT 2006

    210 eMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValA 227
        ||||||||||||||||||||||||||||||| |:::|||||||||:::|||
   2007 CATGGAAAATCTTTTCCTATATGTAATAAGTATTTAGAGTTCATGGTAG 2056

    227 laGluAspTyrMetIleValTyrLeuAsnGlyAlaThr.ProArgArgAr 243
        ||||||||||||||||||||||||||||||| ||. ||| ||||||||||
   2057 CTGAAGACTATATGATTGTGTACTTGAATGGTGCAACCCCCAAGAAGGAG 2106

    243 gMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgA 260
        |||||||||:::|||||||:::|||||||-|| ||||||||||||||||
   2107 GATGCCAGGGCTAGGCTGGATAAAAAAATGCTACCAGATGATTGACAGAC 2156

    260 rgLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpPhe 276
        |||||||||| ||||||||||||:::||| ||  |||||||.|||||||
   2157 CGTTGACGAAAAAATTTAAATCATTTATAATTTCATTCATTTTGTTTT 2206

    277 IleArgThrValLeuAlaIleSerArgProPheIleSerValLysPheIl 293
        |||||||||:::|||||||:::::|||||||| ||||||   ||||||
   2207 ATCAGAACAATCCTTGCTGTGACACGACCTTTTATAAGTTCAAAATTCAG 2256

    293 eAsnLysIleGlnTyrValHisSerLeuGluAspLeuGluGlnLeuIleP 310
        :::'' ::::'|||'|:::|.||||:::::::|||:::   |||||||
```

```
   2257 CAGTAAAATTAAATATGTCAATAGCTTATCAGAACTCAGTGGGCTGATCC 2306

    310 roMetGluHisValGlnIleProAspCysValLeuGlnTyrGluGluGlu 326
        ||||| '     ||||||||:::  :::::::::  :::|.::
   2307 CAATGGATTGCATCCACATTCCAGACAGCATCATCAAACTGGATGAAGAA 2356

    327 ArgLeuLysAlaArgArgGluSerAlaArgPro..............Gl 348
           |||:::         |||'':|||:::                  ::
   2357 ...CTGAGGGAAGCATCAGAGGCAGCTAAAACTAGCTGCCTTTACAATGA 2406

    338 nProGluPhe.................ValLeuProArgSerGluGluL 364
        :|||||||::                 :::     :::   :::|:::
   2404 TCCAGAAATGTCTTCTATGGAGAAGCATATTGACTTGAAGCTGAAAGAAA 2453

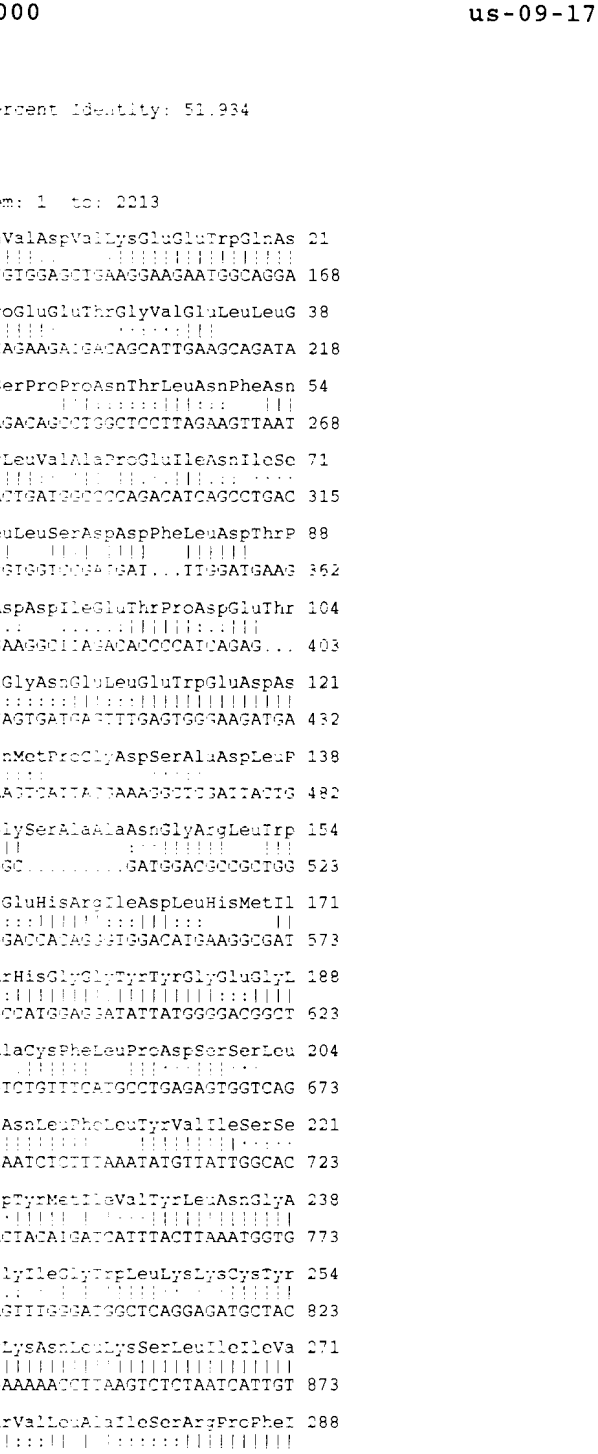
    349 ysProGluValAlaProValGluAsrArgSerAla...LeuValSerGlu 364
        |||||   :::|||   :::|||.......:|||   ||||||
   2454 AGCCTTAGTTGGCCATGCTGGAAGAAGCAGGATGCTTTTCTGGTTCATGGT 2503

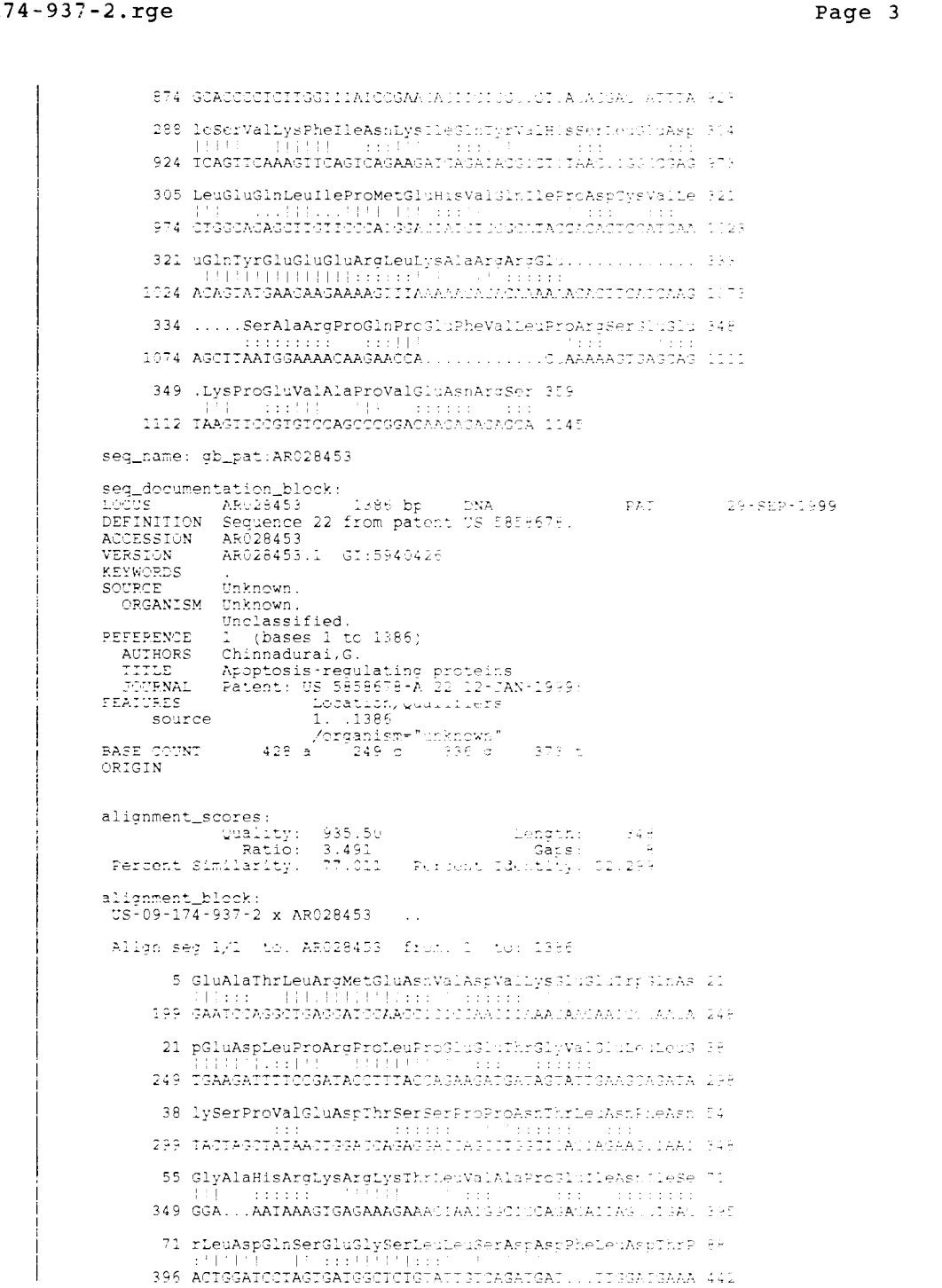
    365 AspGlnGluThr 368
        ...   ||||||
   2504 TCTGTTGAAACA 2515

seq_name: gb_ro:AF035207

seq_documentation_block:
LOCUS       AF035207     2213 bp    mRNA            ROD       25-FEB-1998
DEFINITION  Mus musculus Nip2l mRNA, complete cds.
ACCESSION   AF035207
VERSION     AF035207.1  GI:2911347
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     The mouse Nip2l gene: cloning, sequencing, expression analysis and
            genetic mapping
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Barbosa,M.D.F.S., Detter,J.C. and Kingsmore,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Functional Genomics, CuraGen Corporation,
            555 Long Wharf Drive 11th floor, New Haven, CT 06511, USA
FEATURES             Location/Qualifiers
     source          1..2213
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="between D9Mit39 and D9Mit1"
     gene            1..2213
                     /gene="Nip2l"
     CDS             134..1114
                     /gene="Nip2l"
                     /note="associates with motifs of both the adenovirus E1B
                     19 kDa and Bcl-2 protein"
                     /codon_start=1
                     /product="NIP2l"
                     /protein_id="AAC04329.1"
                     /db_xref="GI:2911348"
                     /translation="MEGVELKEEWQDEDFPIPLPEDDSIEADTLDGTDFQRQPGSLEV
                     NGNKVPKKLMAPDSILDPGFGSLWSDDLDEAGFVDLFRLDTPSENSLEFEWEDDLP
                     KPKTTEVIPKGSITEYTAIEEKGDGPPWPMFRIGPQDHRVDMKAIFPYKKVISHGGYY
                     GDGLNAIVVFAVCPMFESQFNYFYLMDNLFKYVICTLEIIVAPNVMIINLNGATTRR
                     KMPSLGWLPPCYQLIQRLRKNLKSLIIVHPSWFIRILLAVTRPFISSKFSQKIRYVF
                     NLAELAELVPMETVPEFFIPLSYEFEKFPFPQFEVDSELNTPLPFFKPL"
BASE COUNT      637 a    400 c    571 g    605 t
ORIGIN


alignment_scores:
            Quality:  948.50            Length:    362
              Ratio:  3.424               Gaps:      8
```


Percent Similarity: 76.519 Percent Identity: 51.934
alignment_block:
US-09-174-937-2 x AF035207 ..
Align seg 1/1 to: AF035207 from: 1 to: 2213


seq_name: gb_pat:AR028453
seq_documentation_block:
LOCUS AR028453 1386 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 22 from patent US 5858678.
ACCESSION AR028453
VERSION AR028453.1 GI:5940426
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1386)
AUTHORS Chinnadurai,G.
TITLE Apoptosis-regulating proteins
JOURNAL Patent: US 5858678-A 22 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..1386
/organism="Unknown"
BASE COUNT 428 a 249 c 336 g 373 t
ORIGIN
alignment_scores:
Quality: 935.50 Length: 348
Ratio: 3.491 Gaps: 5
Percent Similarity: 77.011 Percent Identity: 52.299
alignment_block:
US-09-174-937-2 x AR028453 ..
Align seg 1/1 to: AR028453 from: 1 to: 1386



```
seq_name: gb_pr3:HSU15173

seq_documentation_block.
LOCUS       HSU15173     2382 bp    mRNA            PRI       02-FEB-1999
DEFINITION  Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2)
            mRNA, complete cds.
ACCESSION   U15173
VERSION     U15173.1  GI:558843
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 212 to 1156)
  AUTHORS   Boyd,J.M., Malstrom,S., Subramanian,T., Venkatesh,L.K.,
            Schaeper,U., Elangovan,B., D'Sa-Eipper,C. and Chinnadurai,G.
  TITLE     Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
            of cellular proteins
  JOURNAL   Cell 79, 341-351 (1994)
  MEDLINE   95042730
REFERENCE   2  (bases 1 to 2382)
  AUTHORS   Boyd,J M
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1994) [illegible] Boyd, Institute for Molecular
            Virology, St. Louis University Health Sciences Center, 3681 Park
            Avenue, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .2382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="B-cell, S.Elledge; B-cell, J.Ambrus; fetal
                     brain, M.Green"
                     /cell_type="B cells; brain"
                     /dev_stage="fetus (brain)"
     gene            1. .2382
                     /gene="BNIP2"
     CDS             212. .1156
                     /gene="BNIP2"
                     /codon_start=1
                     /product="BCL2/[illegible] adenovirus E1B 19kD-interacting
                     protein 2"
                     /protein_id="AAC00021.1"
                     /db_xref="GI:558844"
                     /translation="MEGVELKEEWQDEDFPIPLPEDDSIEAEILAITGEDQPGSLEV
                     NGNKVPKKLMAPDISITDPSDGSVLSDDLDESSEITDFMSDGPSNSNFEWEDDLP
                     KPKTTEVIRKGSITEYTAAEEPEDGPEWPMEPIGFQDHPVDMRAIFPEKRVISHGGYY
                     GDGLNAIVVPAVCPMPFSQPNYPYLMDNLPKYVTGTIPIIVAPNYMIVVLNGATTPR
                     KMPSLGWLRKCYQQLDRLPFNLKSIIIVHPSWFPRIIAVIKPFISSKTSQKIRYVF
                     NLAELAELVPMPYVTIPECTKIVDQELNGKQPEPKNEQ"
BASE COUNT      734 a    393 c    511 g    744 t
ORIGIN


alignment_scores:
           Quality:  935.50            Length:    348
             Ratio:  3.491               Gaps:      8
 Percent Similarity:  77.011   Percent Identity: 52.299

alignment_block
 US-09-174-937-2 x HSU15173   ..

 Align seg 1/1  to: HSU15173  from: 1  to: 2382

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
         ||||:::  |||||||||||||:::|| ::::::|||||||||'|| ||||
     197 GAATCCAGGCTGAGGATGGAAGGTGTGGAACTTAAAGAAGAATGGCAAGA 246

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuG 38
         |||||||...|||  ||||||||||| '::  :::::: |
     247 TGAAGATTTTCCGATACCTTTACCAGAAGATGATAGTATTGAAGCAGATA 296

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
            :::          ::::::  |||:::::::::||||:  ||
     297 TACTAGCTATAACTGGACCAGAGGACCAGCCTGGCTCACTAGAAGTTAAT 346

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
         |||  ::::::  ||||||  || :::||||||:::||:::::::::
     347 GGA ..AATAAAGTGAGAAAGAAACTAATGGCTCCAGACATTAGCCTGAC 394

      71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
         :||||||  |||...||||||:::| ||||||||||  '''''
     394 ACTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT  TTGGATGAAA 44
```

```
     88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104
441 GTGGGGAGATTCACTTA.....GATGGCTTAGACACACCGTCAGAG .. 481

    105 AspSerLeuGluPheLeuGlyAsnGlyAsnGlyLeuGluTrpGluAspAs 121
482 ........ .. AATAGTAATGAGTTTGAGTGGGAAGATGA 510

    121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
511 TCTTCCA... . . ..AAACCTAAGACTACTGAAGTAA 539

    138 heGlyAspGlyThrThrGluAspGlySerAlaAla ..... . Asn 150
540 TTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGAAGAT 589

    151 GlyArgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAs 167
590 GGACGACCTCTGGCGTATGTTCAGGATTGGAGAACAGGACCACAGGGTTGA 639

    167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrT 184
640 TATGAAGGCAATTGAACCCTATAAAAAAGTTATCAGCCATGGGGGATATT 689

    184 yrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuPro 200
690 ATGGGCATGGATTAAATGCCATTGTTGTATTTGCTGTCTGTTTCATGCCT 739

    201 AspSerSerLeuProAspTyrHisTyrIleMetGluAsnLeuPheLeuTy 217
740 GAAAGTAGTCAGCCTAACTATAGATACCTGATGGACAATCTTTTTAAATA 789

    217 rValIleSerSerLeuGluLeuLeuValAlaGluAspTyrMetIleValT 234
790 TGTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAACTACATGATAGTTT 839

    234 yrLeuAsnGlyAlaThrProArgArgArgMetProGlyIleGlyTrpLeu 250
840 ATTTAAATGGTGCAACAACTCGAAGAAAAATGCCCAGTCTGGGATGGCTC 889

    251 LysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLysSe 267
890 AGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAAAAATCTAAAATC 939

    267 rLeuIleIleValHisProSerTrpPheIleArgThrValLeuAlaIleS 284
940 CCTAATCATTGTACATCCTTCTTGGTTTATCAGAACACTTCTGGCTGTTA 989

    284 erArgProPheIleSerValLysPheIleAsnLysIleGlnTyrValHis 300
990 CAAGACCATTTATTAGCTCGAAATTCAGCCAAAAAATTAGATACGTGTTT 1039

    301 SerLeuGluAspLeuGluGlnLeuIleProMetGluHisValGlnIlePr 317
1040 AATTTGGCAGAACTAGCAGAACTTGTCCCCATGGAATACGTTGGCATACC 1089

    317 oAspCysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArgGluS 334
1090 AGAATGCATAAAACAAGTTGATCAAGAA...CTTAATGGAAAACAAGATG 1136

    334 erAlaArgProGlnProGluPheValLeuProArgSerGluGlu 348
1137 AA..... . . .. CCGAAAAATGAACAG 1153

seq_name: gb_htg3:AC011488

seq_documentation_block:
LOCUS       AC011488   179301 bp    DNA             HTG       07-OCT-1999
DEFINITION  Homo sapiens chromosome 19 clone CIT978SKB_171N13, *** SEQUENCING
            IN PROGRESS ***, 54 unordered pieces.
ACCESSION   AC011488
VERSION     AC011488.1  GI:6015264
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179301)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     www.jgi.doe.gov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 54 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1017 bp in length
            *                   gap of unknown length
            *     1008     1936: contig of 929 bp in length
            *                   gap of unknown length
            *     1937     2810: contig of 874 bp in length
            *                   gap of unknown length
            *     2811     3791: contig of 981 bp in length
            *                   gap of unknown length
            *     3792     4639: contig of 848 bp in length
            *                   gap of unknown length
            *     4640     5254: contig of 615 bp in length
            *                   gap of unknown length
            *     5255     5713: contig of 459 bp in length
            *                   gap of unknown length
            *     5714     6345: contig of 632 bp in length
            *                   gap of unknown length
            *     6346     6971: contig of 626 bp in length
            *                   gap of unknown length
            *     6972     7837: contig of 866 bp in length
            *                   gap of unknown length
            *     7838     8339: contig of 502 bp in length
            *                   gap of unknown length
            *     8340     8880: contig of 541 bp in length
            *                   gap of unknown length
            *     8881     9498: contig of 618 bp in length
            *                   gap of unknown length
            *     9499    10379: contig of 881 bp in length
            *                   gap of unknown length
            *    10380    10992: contig of 613 bp in length
            *                   gap of unknown length
            *    10993    11901: contig of 909 bp in length
            *                   gap of unknown length
            *    11902    13352: contig of 1451 bp in length
            *                   gap of unknown length
            *    13353    13947: contig of 595 bp in length
            *                   gap of unknown length
            *    13948    15016: contig of 1069 bp in length
            *                   gap of unknown length
            *    15017    15614: contig of 598 bp in length
            *                   gap of unknown length
            *    15615    16687: contig of 1073 bp in length
            *                   gap of unknown length
            *    16688    18152: contig of 1465 bp in length
            *                   gap of unknown length
            *    18153    19221: contig of 1069 bp in length
            *                   gap of unknown length
            *    19222    20477: contig of 1256 bp in length
            *                   gap of unknown length
            *    20478    21388: contig of 911 bp in length
            *                   gap of unknown length
```

```
            *    21999    22059: contig of 670 bp in length
            *                    gap of unknown length
            *    22059    23154: contig of 1096 bp in length
            *                    gap of unknown length
            *    23155    23925: contig of 771 bp in length
            *                    gap of unknown length
            *    23926    24290: contig of 365 bp in length
            *                    gap of unknown length
            *    24291    24907: contig of 617 bp in length
            *                    gap of unknown length
            *    24908    25839: contig of 932 bp in length
            *                    gap of unknown length
            *    25840    27116: contig of 1277 bp in length
            *                    gap of unknown length
            *    27117    27914: contig of 798 bp in length
            *                    gap of unknown length
            *    27905    28519: contig of 615 bp in length
            *                    gap of unknown length
            *    28520    28930: contig of 411 bp in length
            *                    gap of unknown length
            *    28931    30384: contig of 1454 bp in length
            *                    gap of unknown length
            *    30385    31245: contig of 861 bp in length
            *                    gap of unknown length
            *    31246    32148: contig of 903 bp in length
            *                    gap of unknown length
            *    32149    33058: contig of 910 bp in length
            *                    gap of unknown length
            *    33059    34020: contig of 962 bp in length
            *                    gap of unknown length
            *    34021    34782: contig of 762 bp in length
            *                    gap of unknown length
            *    34783    36041: contig of 1259 bp in length
            *                    gap of unknown length
            *    36042    37033: contig of 992 bp in length
            *                    gap of unknown length
            *    37034    37565: contig of 532 bp in length
            *                    gap of unknown length
            *    37566    38344: contig of 779 bp in length
            *                    gap of unknown length
            *    38345    39359: contig of 1015 bp in length
            *                    gap of unknown length
            *    39360    41242: contig of 1883 bp in length
            *                    gap of unknown length
            *    41243    41839: contig of 597 bp in length
            *                    gap of unknown length
            *    41840    42836: contig of 997 bp in length
            *                    gap of unknown length
            *    42837    43913: contig of 1077 bp in length
            *                    gap of unknown length
            *    43914    55087: contig of 11174 bp in length
            *                    gap of unknown length
            *    55088    81867: contig of 26780 bp in length
            *                    gap of unknown length
            *    81868   112600: contig of 30733 bp in length
            *                    gap of unknown length
            *   112601   179301: contig of 66701 bp in length.
FEATURES             Location/Qualifiers
     source          1..179301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CIT978SKB_171N13"
BASE COUNT    42023 a  47125 c  47470 g  42598 t     85 others
ORIGIN


alignment_scores:
            Quality:  487.00              Length:   1398
              Ratio:  2.231                 Gaps:      5
 Percent Similarity:  15.451   Percent Identity:  15.165

alignment_block:
```

```
US-09-174-937-2 x AC011488/rev   ..

Align seg 1/1  to reverse of: AC011488  from: 1  to: 179301

     45 SerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLysArgLy 61
        ......||||||||||||||||||||||||||||||||||||||||||||
 124307 GCAGCTCCTCCCAACACGCTAAATTTCAACGGAGCCGCATCGTAACGGAA 124258

     61 sThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerGluGlyS 78
        |||||||||||||||||||||||||||||||||||||||||||||||||
 124257 GACGCTGGTGGCCCCAGAGATCAACATTTCTCTGGATCAGAGTGAGGGGT 124208

     78 erLeuLeuSerAspAspPheLeuAspThrProAspAspLeuAspIleAsn 94
        |||||||||||||||||||||||||||||||||||||||||||||||||
 124207 CCCTGCTGTCCGATGACTTCTTGGATACCCCTGATGACCTGGATATTAAC 124158

     95 ValAspAspIleGluThrProAspGluThrAspSerLeuGluPheLeuGl 111
        |||||||||||||||||||||||||||||||||||||||||||||||||
 124157 GTGGATGACATCGAGACCCCCGATGAGACCGACTCGCTGGAGTTCCTGGG 124108

    111 yAsnGlyAsnGluLeuGluTrp.Glu..................... 119
        |||||||||||||||||||| :::
 124107 GAATGGCAACGAACTGGAGTGGGAAGGTAAAGTTCAGGGTCTCTCTCCCC 124058

    119 .. ..... ..... ..... ..... 119

 124057 CCTGCTGGAGCCCACCCCCCCCACCCCACCTTTCCGTCTCTGGATTCCCA 124008

    119 . . ..... 119

 124007 TAGGCTCAGAGAGTCACAAGTGGCCCACCCCCTTAAGCAGTCTAGCCCT 123958

    119 ..... 119

 123957 AAACCCAGGAGATCAAGACTGCAGCGAGACGTGATGATGCCACCGCACTC 123908

    119 ..... 119

 123907 CAGCCTGGACAACAGAGTGAGACCCTGTCTCAAAAATAAAATTTTTAAAA 123858

    119 ..... 119

 123857 AAGAGAGAGGTGGCTGGGCGCAGTGCCTCATGCTGTAATCCTAGCACTT 123808

    119 . ..... 119

 123807 TGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGATCGAGACCATC 123758

    119 ..... 119

 123757 CTGGCTGACACAGTGAAACCCCGTCTCTACTAAAATACAAAAAATTAGCC 123708

    119 ..... . ..... 119

 123707 AGGCATGGTGGCGGGCACCTGAATCCCAGCTACTCAGGAGGCTGAGGCA 123658

    119 ..... ..... ..... 119

 123657 GGAGAACGGTGTGAACCCAGGAGGCGGAGCTTGCGGTGAGCCAAGATTGT 123608

    119 ..... . .... ..... 119

 123607 GCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAA 123558

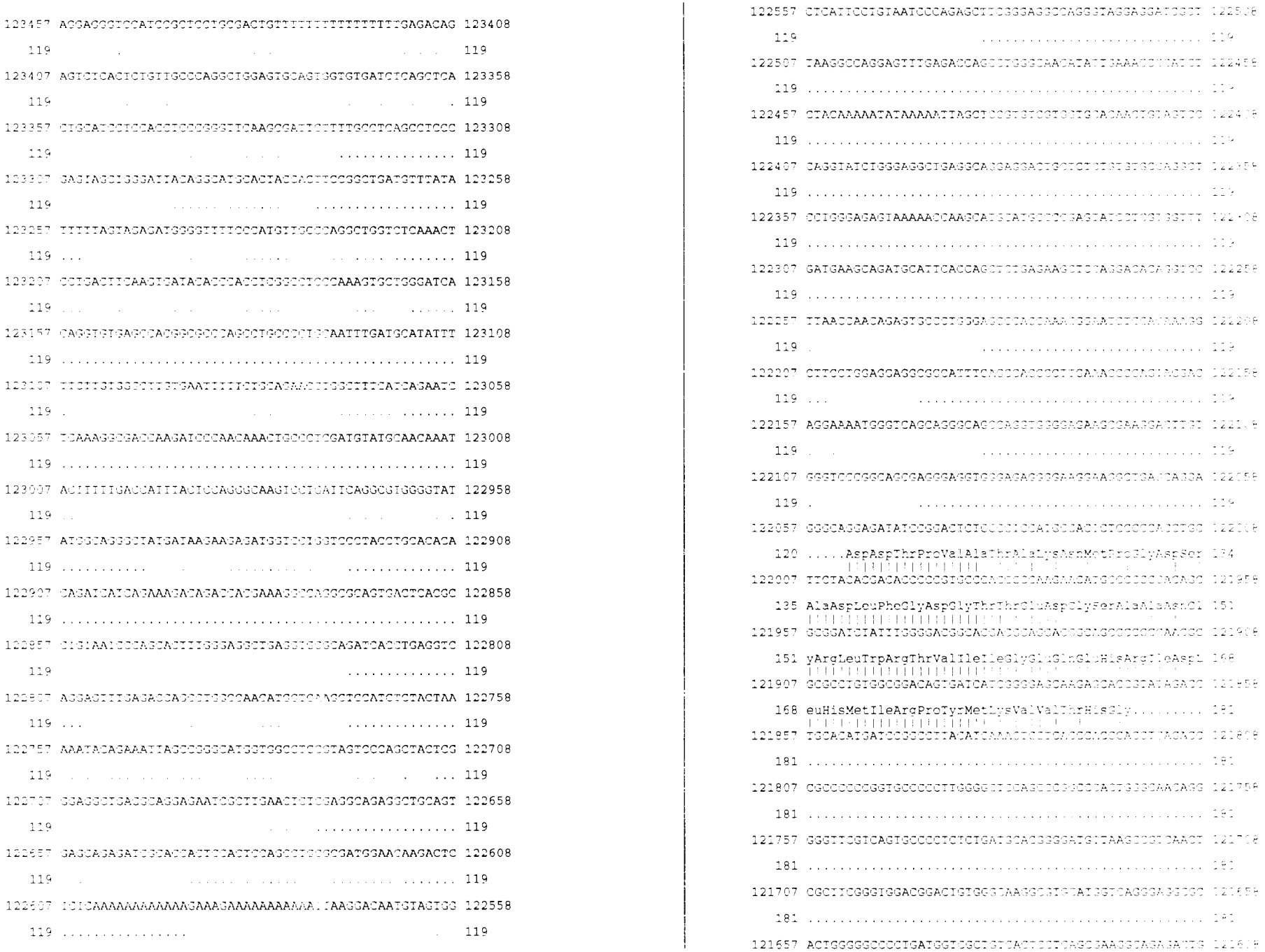
    119 ..... ..... 119

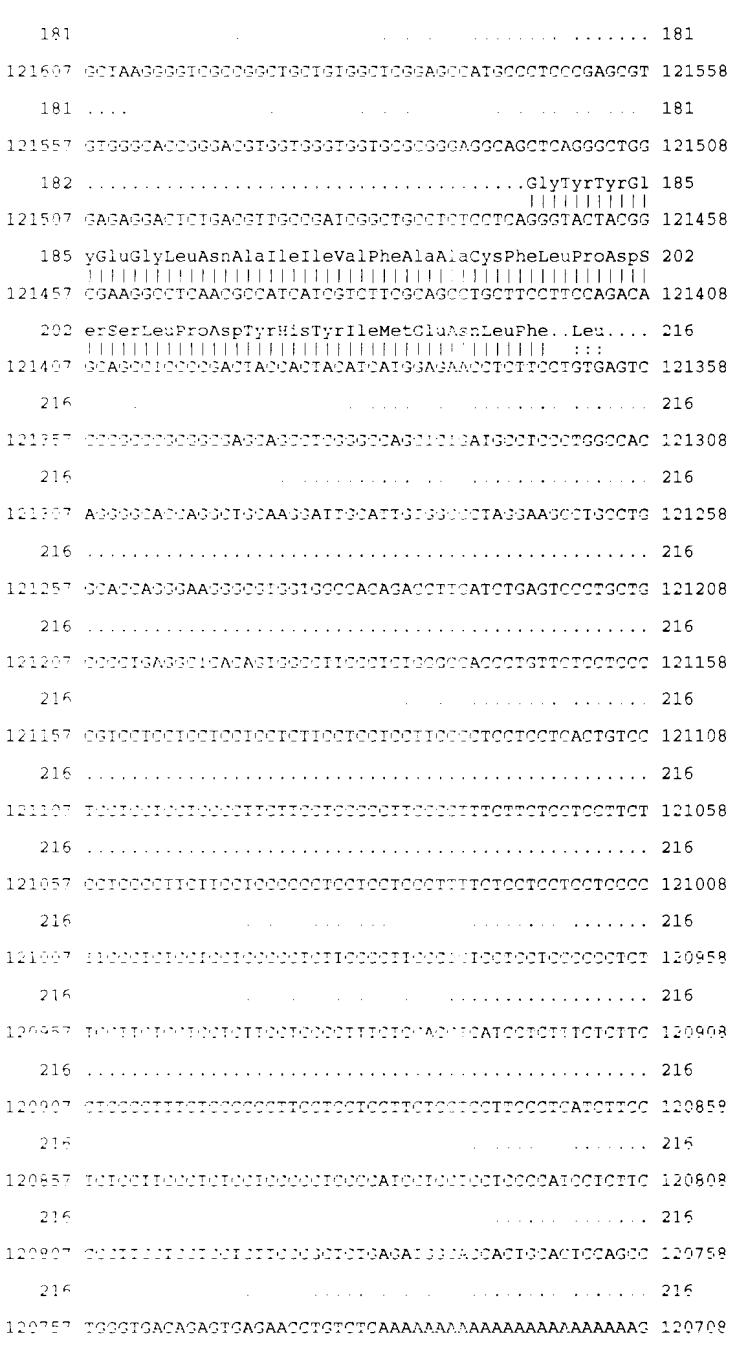
 123557 AAAAAAAAGAGAGAGAGGTTGGTGAATGGGTACCAACATACAGTTAGA 123508

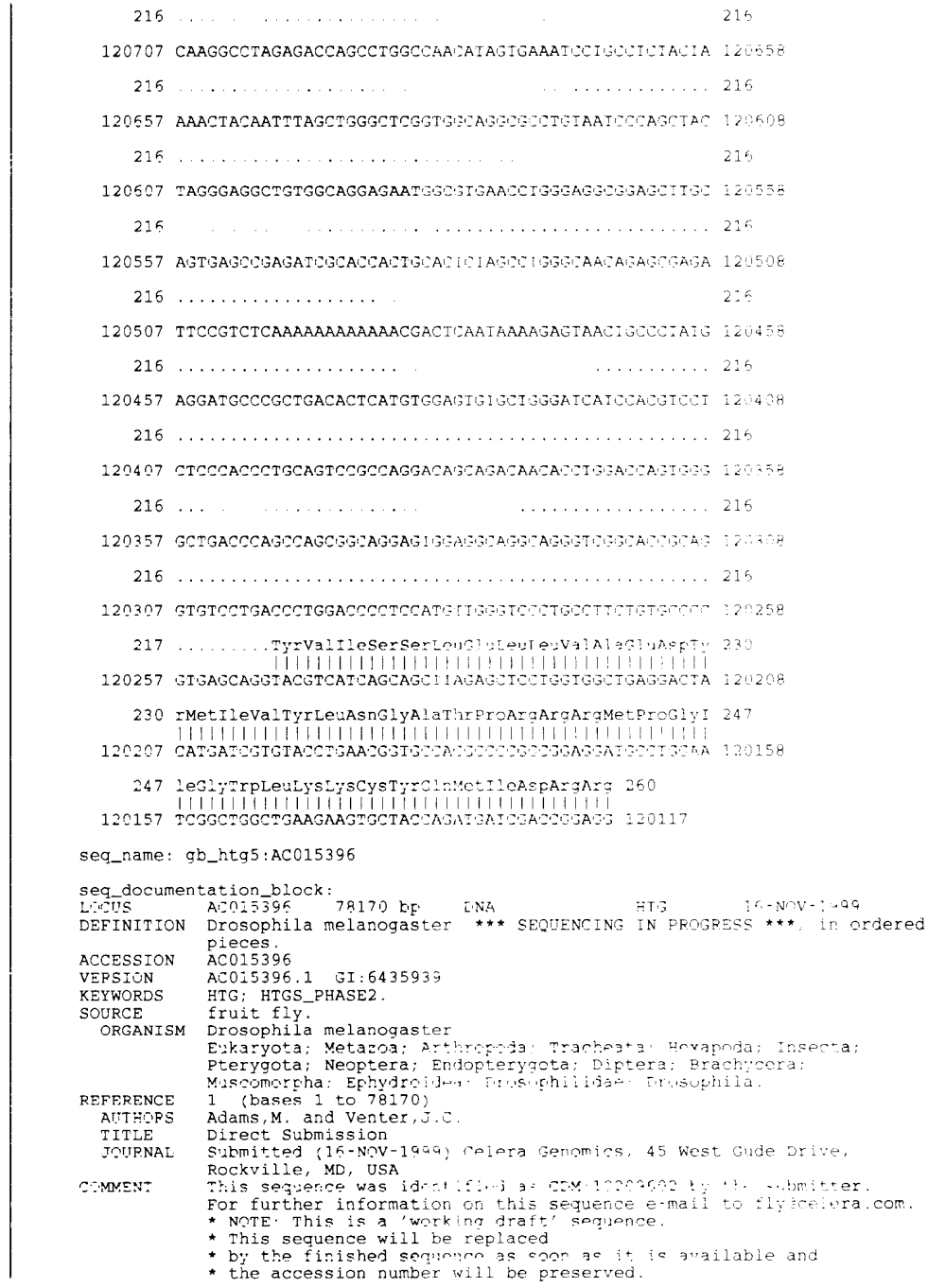
    119 ..... ..... ..... 119

 123507 CAGAAGGAATAAGTTCTATTGTTCGATAGCAGAATAGGAGCGTACCAGC 123458

    119 ..... 119
```







```
seq_name: gb_htg5:AC015396

seq_documentation_block:
LOCUS       AC015396    78170 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster  *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015396
VERSION     AC015396.1  GI:6435939
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78170)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as GSM:10003602 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
```

```
FEATURES             Location/Qualifiers
     source          1. .78170
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21762 a  17119 c  17312 g  21978 t
ORIGIN


alignment_scores:
            Quality:  436.00              Length:    383
              Ratio:  1.938                 Gaps:     15
 Percent Similarity:  58.747   Percent Identity:  32.115

alignment_block:
 US-09-174-937-2 x AC015396/rev  ..

 Align seg 1/1  to reverse of. AC015396  from: 1  to: 78170

       9 LeuArgMetGluAsnValAspValLysGluGluTrpGlnAspGluAspLe 24
         ||||||  ||||||  |||:::  ||||||  ||||||  ||||::
    9180 CTAAGGGCCAACAGTCCGGATATCCTGAGCAGCGAGTCAGATGCGGATGT 9131

      24 uProArgProLeuProGluGluThrGlyValGluLeuLeuGlySerProV 41
         .  ...  |||          ||||||||||          |||
    9130 GTCCCAGTACTTG .    . GCCGCCGTGGAG  . . ....TCTAACT 9099

      41 alGluAspThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHis 57
         . .  ||||      ||||||            ||||||
    9098 TTCAGTCCGTTTCCCTGATGCCCAATGGC       ...AATGGCAAGAAA 9058

      58 ArgLysArgLysThr...LeuValAlaProGluIleAsnIleSerLeuAs 73
         ||||||:.|||  |||  ...  ||          ||||||||
    9057 ICCAAACGAICGACTGCACICCCTGGTGGCGAGGACATATCTAGCCTGGA 9008

      73 pGluSerGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspA 90
         ||||              ||.|||...     ...|||      ||||
    9007 TTCC..             ATCTCCAAICACAGCTTCGACGATGACGAGG 8973

      90 spLeuAspIleAsn. ...  ......................ValAsp 96
         ||..       |||                            ||||||
    8972 ACATTGACGACAATTTGGCCTTTCTAAGTTCATCGAGTTCCTTGATTGAT 8923

      97 AspIleGluThrProAspGluThrAspSerLeuGlu  ...  .....Ph 109
         ||||||     |||:.  |||...  |||             ||
    8922 GGCTTAGACGGAGAGGATGACGACGACGACTGTGAGGGTCCGGAACTGCT 8873

     109 eLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProValAlaT 126
         :  ...  .....|||||  .....  ...  ...|||
    8872 GCCCGACCACGACGACGACTTGGAACTGGACCTCGAGTTCGGTTGGCCA 8823

     126 hrAlaLysAsnMetProGlyAspSerAlaAspLeuPheGlyAspGlyThr 142
         |||||  ||.  |||..  ||||...  ...  ..:|||
    8822 CCAGCCCAGG  GCAAATGTTCCGGGCATCGCCTAGGGCTGGCTCAACT 8776

     143 ThrGluAspGlySerAlaAla.. ....      AsnGlyArgLeuTrpAr 155
         .  .  ..||||||              ::|||  |||..
    8775 CTGCCACAGTATACGGCTGCTGAAGAGAGACGTGATTCGCGGAATTGGCA 8726

     155 gThrValIleIleGlyGluGlnGluHisArg...IleAspLeuHisMetI 171
         :  :::  ...  ||  ||||||  .  :|
    8725 GAAGATCACTTTACCGGATGGACGTACCAGGCAGATCGACATGCGGGTCA 8676

     171 leArgProTyrMetLysValValThrHisGlyGlyTyrTyr...GlyGlu 186
         |:::|||||  :.:|||.....|||| | |||||      ..:
    8675 TAGAGCCCTACAAGCGTGTGCTGTCCCACGGCGGCTATCTCAAATCTGGC 8626

     187 GlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSerSe 203
         |:|  ||||||||  ...  :|||  ||||||  ||||||||  ||
    8625 GGTCAGAACGCCATTGTCATCTTTTGCGCCTGTCACCTGCCGGATAGGTC 8576

     203 rLeuProAspTyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleS 220
         |  ||||||  |||:::|||:::|| | ||||||||||||||::::
    8575 GAGGGCCGACTACAGCTACGTGATGGACAATTTGTTCTGTACGTGGTCA 8526

     220 erSerLeuGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsn 236
         ::::|||||  |||  |||  ||||||  ||||||  ||||||  |||
    8525 ACACCCTGGAAACAGTTGGTCACTCATGACATCTTGCAAATGTACTCCAAT 8476

     237 GlyAlaThrProArgArgArgMetProGlyIleGlyTrpLeuLysLys 252
         |||::::::  ||||||:::::  :::  |  |
    8475 GGCGGCTCCAATAGGCGTAATGTGCCACCTTTTCCCTGGCTTAAACGGTG 8426

     252 ..................................................... 252

    8425 ATTTGTTTGGTTTGTTTTTTAA[illegible]TA 8376

     253 .       .CysTyrGlnMetIleAspArgArgLeuArgLysAspLeuLy 266
                 ||||||||:::::|  .  ...  |||  :::
    8375 TCCTTTAGGTGTTATCAGCTGCTTCATCCCCGATTGCGCAAAACACTCTAA 8326

     266 sSerLeuIleIleValHisProSerTrpPheIleArgThrValLysAlaI 283
         |  :::  :::|||||||||:::::::::|  :::::::::::  :
    8325 GCACATGTATTTGGTGCATCCGACATTCTGGATAAAATCGCTGGTTTGTGA 8276

     283 leSerArgProPhe..................................... 287
         :::::|||||||||
    8275 TGGCGCGGCCTTTTGTGAGGTAAGTAATAATAATCATTGTCAAATGTAGA 8226

     288 ..                          Ile..SerValLysPheIleAspLys 295
                                     :::  |:::  :  :::  :::
    8225 TCTCATATAAATACTGTTTATTTTTTTAGGTCAAATTTTTCTCAAAG 8176

     296 IleGlnTyrValHisSerLeuGluAspLeuGluGlnLeuIlePr MetGl 312
         ...  ||||||  ||||||||  :::  |  :::  :::  :::::
    8175 CTGGTCTACGTAAAGTCCCTGGAGGAACTGGGCATGCACGTGGTCTGGA 8126

     312 uHisValGlnIleProAspCysValLeuGlnTyrGluGluGluArg 327
         |  :::  ||||||:::  |  :::  :::  :::
    8125 GAAGGCAGCCATTCCGGAGAAGCTCAACCAATACGACCCCAAGCTG 8081

seq_name: gb_pri:HSRHOGAPX

seq_documentation_block:
LOCUS       HSRHOGAPX    1430 bp    mRNA            1-1       11-MAR-1997
DEFINITION  H.sapiens rhoGAP protein.
ACCESSION   Z23024
VERSION     Z23024.1  GI:812211
KEYWORDS    rhoGAP protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   Hall,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1993) Alan Hall Dr., Cell and Molecular Biology,
            Institute of Cancer Research, Chester Beatty Laboratories, 239
            Fulham Road, London, SW3 6JB, England
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="Male"
                     /tissue_type="fibrosarcoma"
                     /cell_line="HT1080"
     5'UTR           1. .98
     CDS             99. .1418
                     /codon_start=1
                     /product="rhoGAP protein"
```

```
                     /protein_id="CAA80560.1"
                     /db_xref="GI:312212"
                     /db_xref="SWISS-PROT:Q07960"
                     /translation="MDPLSELQDDLTLDDTSEALNQLKLASIDEKNWPSDEMPDFPKS
                     DDSKSSSPELVTHLKWDDPYYDIARHQIVEVAGDDKYGRKIIVFSACPMPPSHQLDHS
                     KLLGYLKHTLDQYVESDYTLLYLHHGLTSDNKPSLSWLRDAYREFDRKYKKNIKALYI
                     VHPTMFIKTLLILFKPLISFKFGQKIFYVNYLSELSEHVKLEQLGIPRQVLKYDDFLK
                     STQKSPATAPKPMPPPPPLPNQQFGVSLQHLQEKNPEQEPIPIVLRETVAYLQAHALT
                     TEGIFPPSANTQVVPEVQQKYNMGLPVDFDQYNELHLPAVILKTFLPELPEPLLTFDL
                     YPHVVGFLNIDESQFVPATLQVLQTLPEENYQVLPFLTAFLVQISAHSDQNKMTNTNL
                     AVVFGPNLLWAKDAAITLKAINPINTFTKFLLDHQGELFPSPDPSGL"
BASE COUNT      332 a    453 c    365 g    280 t
ORIGIN


alignment_scores:
            Quality:  285.50              Length:    294
              Ratio:  1.807                 Gaps:     10
 Percent Similarity:  53.741   Percent Identity:  28.231

alignment_block:
 US-09-174-937-2 x HSRHOGAPX   ..

 Align seq 1/1  to: HSRHOGAPX  from: 1  to: 1430

      89 AspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThrAs 105
         ||| ||:::         |||||::         |||  :::::
     102 GATGGGCTCTCAGAGGTGCAGGATGATCTGACCTTGGATGACACCAGCGA 151

     105 pSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspT 122
         :::|||          |||:::|'|:::     :::
     152 GGCTCTG.................AACCAGCTGAAGCTGGCCTCC.... 179

     122 hrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuPhe 138
                  ||||||  |||::::||...      |||
     180         ATCGATGAGAAGAACTGGCCCTCGGATGAAATGCCTGACTTC 221

     139 GlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeu..... 153
                 ::   ||| .  .       : |||
     222 .....GCTAATTCATATGACTCCAAAAGCCATGTCCCCGGAACTTGTCAC 265

     154 ..........TrpArgThrValIleIleGlyGluGlnGluHisArgIleA 167
                   |||
     266 ATACCTGAATGGGATGAC.......... .  ...  ..   . .   284

     167 spLeuHisMetIleArgProTyrMetLysValValThrHis     . .   180
                      ||||||      ::::::   |||
     285      . . . . .  CCATACTATGACATCCCCGGGCACCAGATCGTG 317

     181             GlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleVa 193
                        ||||||         :::   ||||||||
     318 GAGGTGGCAGGAGATGACAAGTATGGG.        CGGAAGATCATTGT 358

     193 lPheAlaAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrI 210
         ||||:::||||||   ...|||   |||         |||...
     359 GTTTAGTGCCTGTCGAATGCCCCCCAGCCACCAGCTCGACCAC.. .... 401

     210 leMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuVal 226
           :::::.|||...  |||          :::|||:::      |||
     402 .....AGCAAGCTCCTGGGGTACCTGAAGCACACCCTGGACCAGTACGTG 446

     227 AlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgAr 243
         :::|||||         ||:||:  .  .  .          ::
     447 GAGAGTGACTACACACTTCTGTATCTGCACCACGGCCTGACCAGCGACAA 496

     243 gMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgA 260
         :   |||.........||||||:::   ...|| ...   ...||||||:
     497 CAAGCCCTCCCTCAGCTGGCTCCGTGATGCCTACCGGGAGTTTGACCGCA 546

     260 rgLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpPhe 276
         ::   :::|||||...|||:::||   ||||:|||||||:::   |||
     547 AGTACAAGAAGAACATCAAGCCCTTGTACATCCTGCATCCAACCATGTTC 596

     277 IleArgThrValLeuAlaIleSerArgProPheIleSerValLysPheIl 293
         |||.:|||:::|||   ...     :|||:::|||||   ||||||
     597 ATCAAAACTCTGCTCATCCTCTTCAAGCCCCICATCAGCIICAAGITCGG 646

     293 eAsnLysIleGlnTyrValHisSerLeuGluAspLeuGluGlnLeuIleP 310
          :::|||||   ||||||:::   |||:::::|||:::::   :::
     647 GCAGAAGATCTTCTATGTGAATTACCTGAGGGAGCTGAGCGAGCACGTGA 696

     310 roMetGluHisValGlnIleProAspCysValLeuGlnTyrGluGluGlu 326
          :::|||:::::   ||||||      ||||||...|||   ::::::
     697 AGCTGGAGCAGCTGGGGATCCCTCGCCAAGTGCTCAAATAT   GACGAC 743

     327 ArgLeuLysAlaArgArgGluSerAlaArgProGlnProGluPheValLe 343
           |||||||:::   ::::::|||          |||:::   :::
     744 TTCCTGAAATCCACACAGAAGAGCCCCGGGACAGCCCCCAAGCCCATGCC 793

     343 uProArg..............     .....  .          ..... 345
          ||||||
     794 CCCACGGCCCCCCCTGCCCAACCAGCAGTTTGGAGTCTCGCTGCAGCACC 843

     346 ..SerGluGluLysProGluValAlaProVal 355
           :::|||:::::|||||      |||:::
     844 TCCAGGAGAAGAATCCAGAGCAGGAGCCCATT 875

seq_name: gb_pr1:HSU02570

seq_documentation_block:
LOCUS       HSU02570     3345 bp    mRNA            PRI       25-JAN-1994
DEFINITION  Human CDC42 GTPase-activating protein mRNA, partial cds.
ACCESSION   U02570
VERSION     U02570.1  GI:450780
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3345)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 268, 26059-26062 (1993)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 3345)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P., Self,A.J., Brill,S., van Erp,H.E.
            and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase Activating Protein for
            Rho-Related Small GTPases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3345)
  AUTHORS   Barfod,E.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1993) Elisabeth T. Barfod, Genetech, Inc., 460
            Point San Bruno Blvd, South San Francisco, CA 94080, USA
COMMENT     On Jan 27, 1994 this sequence version replaced gi:409026.
FEATURES             Location/Qualifiers
     source          1. .3345
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CMK"
     CDS             <1. .1391
                     /codon_start=3
                     /function="GTPase activating protein for the rho family of
                     small GTP-binding proteins that is most effective against
                     CDC42"
                     /product="CDC42 GTPase-activating protein"
                     /protein_id="AAA16142.1"
                     /db_xref="GI:409027"
                     /translation="GSSARAPFLLRFCRAPLSGTGIAMDPLSELQDDLTLDDTSEALN
                     QLKLASIDEKNWPSDEMPDFPKSDDSKSSSPELVTHLKWDDPYYDIARHQIVEVAGDD
                     KYGRKIIVFSACPMPPSHQLDHSKLLGYLKHTLDQYVESDYTLLYLHHGLTSDNKPSL
                     SWLRDAYREFDRKYKKNIKALYIVHPTMFIKTLLILFKPLISFKFGQKIFYVNYLSEL
                     SEHVKLEQLGIPRQVLKYDDFLKSTQKSPATAPKPMPPPPPLPNQQFGVSLQHLQEKN
```

```
                     PEQEPIPIVLRETVAYLQAHALTIEGIFPPSANTQVVPEVQQKYNMGLPVDFDQYNEL
                     HLPAVILKTFLFPLPEPLLIFDLYPHVVGFLNIDESQPVPATLQVLQTLPEENYQVLR
                     FLTAFLVQISAHSDQNKMTNTNLAVVFGPNLLWAKDAAITLKAINPINTFTKFLLDHQ
                     GELFPSPDPSGL"
     misc_feature    753. .785
                     /note="SH3-binding domain; capable of binding the SH3
                     domains of src and p85 alpha; aa residues 251-261"
     misc_feature    849. .926
                     /note="active site box 1; aa residues 283-308"
     misc_feature    999. .1091
                     /note="active site box 2; aa residues 333-363"
     misc_feature    1149. .1277
                     /note="active site box 3; aa residues 383-425"
     polyA_signal    3328. .3333
     polyA_site      3345
                     /note="38 A residues"
BASE COUNT      655 a   1061 c    855 g    774 t
ORIGIN


alignment_scores:
            Quality:  285.50              Length:    294
              Ratio:  1.807                Gaps:     10
 Percent Similarity:  53.741   Percent Identity: 28.231

alignment_block:
 US-09-174-937-2 x HSU02570   ..

 Align seg 1/1  to: HSU02570  from. 1  to. 3345

      89 AspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThrAs 105
         |||   |||:::          ||||||...        |||   .....
      75 GATCGGCTCTCAGAGCTGCAGGATGATCTGACCTTGGATGACACCAGCGA 124

     105 pSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspT 122
         :::.|||                |||:::|||...     :::
     125 GGCTCTG.....    .....   AACCAGCTGAAGCTGGCCTCC.... 152

     122 hrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuPhe 138
                       ||||||   |||:::  |:::         |||
     153 ........ATCGATGAGAAGAACTGGCCCTCGGATGAAATGCCTGACTTC 194

     139 GlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeu..... 153
                 :....|||  :.....     .   :.|||
     195 ......CCCAAGTCAGATGACTCCAAAAGCAGCTCCCCGGAACTTGTCAC 238

     154 .      TrpArgThrValIleIleGlyGluGlnGluHisArgIleA 167
                |||
     239 ACACCTGAAGTGGGATGAC....  ....          ..............  257

     167 spLeuHisMetIleArgProTyrMetLysValValThrHis......... 180
                       ||||||      ::::::   |||
     258 .....  ....      CCATACTATGACATCCCCCGGCACCAGATCGTG 290

     181 ............GlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleVa 193
                        ||||||          ...   ||||||||
     291 GAGGTGGCAGGAGATGACAAGTATGGG        ..CGGAAGATCATTGT 331

     193 lPheAlaAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrI 210
         |||||::||||||   ...|||  |||          |||...
     332 GTTTAGTGGCTGTGGAATGCCCCCAGCCAGCAGCTGGACCAC..... 374

     210 leMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuVal 226
         :.....|||...  |||...     ...|||...        |||
     375 .    AGCAAGCTCCTGGGGTACCTGAAGCAGACCCTGGACCAGTACGTG 419

     227 AlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgAr 243
         ...|||||   .....||||||...  ::.             ::
     420 GAGAGTGACTACACACTTCTGTATCTGCAGCTAGGCCTGACCAGCGACAA 469

     243 gMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgA 260
         .  |||.....|||||:...     |||  .   ::|||||:
     470 CAAGCCCTCCCTCAGCTGGCTCCGTGATGCCTACCGGGAGTTTGACCGCA 519

     260 rgLeuArgLysAsnLeuLysSerLe.IleIleVa.HisProSerTrpPro 27
         ::   :::|'|'|'::::'|::::  '     '     '   :::
     520 AGTACAAGAAGAACATCAAGGCCTTGTACATCGTGCATCAACGCATTTTC 56

     277 IleArgThrValLeuAlaIleSerArgPr.PheIleSerValLysPheIl 29
         |||...|||.::|||   :::  :::  :::  '  '
     570 ATCAAAACTCTGCTCATCCCTCTTCAAGCCCCTCATCAGCTTCAATTTCG 61

     293 eAsnLysIleGlnTyrValHisSerLeuGl.AspLe.GluGlnLe.IleP 31
         :::|||||   |||||:::   ||  :::::  |:::::   :::
     620 GCAGAAGATCTTCTATGTGAATTACCTCAGTGACCTCAACGACAACGTGA 66

     310 roMetGluHisValGlnIleProAspCysValLeuGlnTyrGluGlnGln 32
         :::|||:::::   ||||||'     '  :::  :   ::::::
     670 AGCTGGAGCAGCTGGGGATCCCTGGCCAAGTGTTCAAATAT...AACGAC 71

     327 ArgLeuLysAlaArgArgGluSerAlaArgProGlnProGluPheValIe 34
         |:|||:::   :::::::.          :  :::   :::
     717 TTCCTGAAATCCACACAGAAGAGTCCCGCGACAGCCCCCAAGCCTATTTC 76

     343 uProArg..................................................  34
         |||||:
     767 CCCACGGCCCCCCCTGCCCAACCAGTATTTGGAGTCTTGTTGCATCATC 81

     346 ..SerGluGluLysProGluValAlaProVal 355
         .........:::|||||      '  :::
     817 TCCAGGAGAAGAAATCCAGAGCAGGAGCCCATT 848

seq_name: gb_htg6:AC009556

seq_documentation_block:
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-577G21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-577G21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Burton,N., Beckerle,P., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayle,J., Dorsch,K., Dowar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,J., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,C.D., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Liu,C., Lopez,K., Macdonald,P.,
            Marquis,N., McEwan,P., McCurk,A., McPherson,R., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morton,J., Mudhuchakyt,C.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6011499.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.wi.mit.edu/pub/software/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

```
    Contact: sequence_submission@genome.wi.mit.edu
-------------- Project Information
    Center project name: L1672
    Center clone name: 57_C_21
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-primer-amersham; 5% of reads
    Chemistry: Dye-terminator Big Dye; 95% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 136385 bases at least Q40
    Consensus quality: 156543 bases at least Q30
    Consensus quality: 165534 bases at least Q20
    Insert size: 165000; agarose-fp
    Insert size: 169798; sum-of-contigs
    Quality coverage: 4.2 in Q20 bases; agarose-fp
    Quality coverage: 4.1 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1016: contig of 1016 bp in length
*                   gap of unknown length
*     1017     2036: contig of 1020 bp in length
*                   gap of unknown length
*     2037     3371: contig of 1335 bp in length
*                   gap of unknown length
*     3372     4972: contig of 1601 bp in length
*                   gap of unknown length
*     4973     5981: contig of 1009 bp in length
*                   gap of unknown length
*     5982     8008: contig of 2027 bp in length
*                   gap of unknown length
*     8009     9286: contig of 1278 bp in length
*                   gap of unknown length
*     9287    11243: contig of 1957 bp in length
*                   gap of unknown length
*    11244    14312: contig of 3069 bp in length
*                   gap of unknown length
*    14313    18423: contig of 4111 bp in length
*                   gap of unknown length
*    18424    23065: contig of 4642 bp in length
*                   gap of unknown length
*    23066    30558: contig of 7493 bp in length
*                   gap of unknown length
*    30559    37537: contig of 6979 bp in length
*                   gap of unknown length
*    37538    44831: contig of 7294 bp in length
*                   gap of unknown length
*    44832    50944: contig of 6113 bp in length
*                   gap of unknown length
*    50945    58263: contig of 7319 bp in length
*                   gap of unknown length
*    58264    65216: contig of 6953 bp in length
*                   gap of unknown length
*    65217    72310: contig of 7094 bp in length
*                   gap of unknown length
*    72311    78799: contig of 6489 bp in length
*                   gap of unknown length
*    78800    87607: contig of 8808 bp in length
*                   gap of unknown length
*    87608    98337: contig of 10730 bp in length
*                   gap of unknown length
*    98338   112198: contig of 13861 bp in length
*                   gap of unknown length
*   112199   126330: contig of 14132 bp in length
*                   gap of unknown length
*   126331   140617: contig of 14287 bp in length
*                   gap of unknown length
*   140618   153842: contig of 13225 bp in length
*                   gap of unknown length
*   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51330 a  34914 c  34487 g  49065 t      2 others
ORIGIN


alignment_scores:
           Quality:  279.00            Length:    291
             Ratio:  3.623               Gaps:      1
 Percent Similarity:  26.460   Percent Identity: 24.399

alignment_block:
 US-09-174-937-2 x AC009556/rev   ..

 Align seq 1/1  to reverse of: AC009556  from: 1  to: 169798

    182 GlyTyrTyrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 198
        ||||||||||||:::|||||||||||||||||:|||||||||||| | ||
  13902 GGATACTATGGGGACGGTCTAAATGCCATCATTGTGTTTGCCGCCTGTTT 13853

    198 eLeuProAspSerSerLeuProAspTyrHisTyrIleMetGluAsnLeuP 215
        ||||||||||||||||      ||||||||||:::|||||||||||||'
  13852 TCTGCCAGACAGCAGTCGGGCGGATTACCACTATGTCATGGAAAATCTTT 13803

    215 heLeu.  ..........  .                              214
        |||||
  13802 TCCTGTGAGTGATACTTATGACAAAGTGCTTGAGTTAATTCTCTGTGCA 13753

    216 ..................................................  214

  13752 TGTAATGGCAGATAGAGCACCTTTAATTAAAGAAAATAGATGATGTACAA 13703

    216 ..................                                  214

  13702 TGTTCTAAGCCTTCAGTGAAAAGTGATAAGTTTCTGTCAAGGAATCTT 13653

    216 ..  .  ..  .....  .                              214

  13652 GCTCAAGGGGCTAAAAGCTTTAAAGGAGCTCTTTTGTGGGTAGACTAGAC 13603

    216 ..................  .  ..................  ..... 214

  13602 TGAAACCAAGATGAACAGAATTTTGCTTGCATTTTAAAATTCTCTAACAT 13553

    216  .  .. .... ..................................... 214

  13552 TTCCTTTGGGCAAAGTATTATTTTTGTTCCTTCATAACATGTTCTTAGTC 13503

    216 ..................  .                              214

  13502 AGTTAAGACTGACTATAATCTGATTTCATTAATATCTTGCTTTATTTAAC 13453

    216  .  .. .... .  ....  .                 ..... 214

  13452 CTGAGTTACCCTTAGTATTTGATAACATTGTTGACAGTTTTGAAAGTCAG 13403

    216  .. ..................................... 214

  13402 GATTCTTTCTAACCCATACATTTTTCTAGTGTACCTCCATTTAAAAGATA 13353

    216 ..................................................  214

  13352 GATAGCCATACCGATAAATGAGTAATACTCTCATAGGAATGCTCTTAGGT 13303

    216 ..... .. .. ...  .                 ..... 214
```

```
    13302 AATTGTATTCTTCCCAAGGTAAAGAGGTAATATGCACAATCATTCTTAAA 13253

      216 ...                ...    ...          ......  ...... 215

    13252 TTTCATACAGCCTGTACTTTTGATGTTATGTAATTATGGTGTCAATAGGA 13203

      217 ..                 . .    ....         . . .  TyrValIle 219
                                                        ||||||||
    13202 TGTGGGAACTGAAAACATTCCTCTTCCTTCCTGTGATCAGATATGTAATA 13153

      220 SerSerLeuGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAs 236
          |||:::|||||||||:::|||||||||||||:|:|||||||||||||||
    13152 AGTAGTTTAGAGTTGATGGTAGCTGAAGACTATATGATTGTGTACTTGAA 13103

      236 nGlyAlaThrProArgArgArgMetProGlyIleGlyTrpLeuLysLysC 253
          |||||||||||||||||||||||||||||:::|||||:::|||||||
    13102 TGGTGCAACCCCAAGAAGGAGGATGCCAGGGCTAGGCTGGATGAAGAAAT 13053

      253 ysTyrGlnMetIleAspArgArg 260
          ||||||||||||||||||||||
    13052 GCTACCAGATGATTGACAGACGG 13030

Seq_name  gb_htg6 AC009556

seq_documentation_block:
LOCUS       AC009556   169798 bp    DNA             HTG       30-NOV-1999
DEFINITION  Homo sapiens chromosome 15 clone RP11-57C21 map 15, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC009556
VERSION     AC009556.3  GI:6479158
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-57C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169798)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,T., Baldwin,J., Barna,N., Beckerly,P., Bonn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Ichoczky,J., Liu,C., Locke K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Poy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 30, 1999 this sequence version replaced gi:6114933.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submission@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1672
                Center clone name: 57_C_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156562 bases at least 40
                Consensus quality: 156543 bases at least 50
                Consensus quality: 155534 bases at least 60
                Insert size: 165000; agarose-ip
                Insert size: 169798; sum-of-contigs
                Quality coverage: 4.2 in 620 bases; agarose-ip
                Quality coverage: 4.1 in 620 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1016: contig of 1016 bp in length
            *                   gap of unknown length
            *     1017     2036: contig of 1020 bp in length
            *                   gap of unknown length
            *     2037     3371: contig of 1335 bp in length
            *                   gap of unknown length
            *     3372     4972: contig of 1601 bp in length
            *                   gap of unknown length
            *     4973     5981: contig of 1009 bp in length
            *                   gap of unknown length
            *     5982     8008: contig of 2027 bp in length
            *                   gap of unknown length
            *     8009     9286: contig of 1278 bp in length
            *                   gap of unknown length
            *     9287    11243: contig of 1957 bp in length
            *                   gap of unknown length
            *    11244    14312: contig of 3069 bp in length
            *                   gap of unknown length
            *    14313    18423: contig of 4111 bp in length
            *                   gap of unknown length
            *    18424    23065: contig of 4642 bp in length
            *                   gap of unknown length
            *    23066    30558: contig of 7493 bp in length
            *                   gap of unknown length
            *    30559    37537: contig of 6979 bp in length
            *                   gap of unknown length
            *    37538    44831: contig of 7294 bp in length
            *                   gap of unknown length
            *    44832    50944: contig of 6113 bp in length
            *                   gap of unknown length
            *    50945    58263: contig of 7319 bp in length
            *                   gap of unknown length
            *    58264    65216: contig of 6953 bp in length
            *                   gap of unknown length
            *    65217    72310: contig of 7094 bp in length
            *                   gap of unknown length
            *    72311    78799: contig of 6489 bp in length
            *                   gap of unknown length
            *    78800    87607: contig of 8808 bp in length
            *                   gap of unknown length
            *    87608    98337: contig of 10730 bp in length
            *                   gap of unknown length
            *    98338   112198: contig of 13861 bp in length
            *                   gap of unknown length
            *   112199   126330: contig of 14132 bp in length
            *                   gap of unknown length
            *   126331   140617: contig of 14287 bp in length
            *                   gap of unknown length
            *   140618   153843: contig of 13226 bp in length
            *                   gap of unknown length
            *   153844   169798: contig of 15955 bp in length.
FEATURES             Location/Qualifiers
     source          1..169798
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-57C21"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51230 a  24014 c  24487 g  10065 t      2 others
ORIGIN


alignment_scores.
            Quality:  254.00              Length:     73
              Ratio:  3.791                Gaps:      0
 Percent Similarity:  91.781   Percent Identity: 60.274

alignment_block:
 US-09-174-937-2 x AC009556   ..

 Align seg 1/1  to: AC009556  from: 1  to: 169738

      47 ProProAsnThrLeuAspPheAsnGlyAlaHisArgLysArgLysThrLe 63
         ||||||||||::|||:.....|||   ...||:   ...|||      ||
  119828 CCTCCIAAIICICIGGATCTTAATGACACTCAICCTCGGAGAATCAAGCT 119877

      63 uValAlaProGluIleAsnIleSerLeuAspGlnSerGluGlySerLeuL 80
         ::::| | |:::::|||||:::::|||||| ||  ||||||||||||:::|
  119878 CACAGCCCCAAAIAICAAICTTTCTCTGGACCAAAGTGAAGGATCTATTC 119927

      80 euSerAspAspPheLeuAspThrProAspAspLeuAspIleAsnValAsp 96
         |||||||||||   |||||:::|||||:::::::|||||||||||||||
  119928 TCTCTGATGATAACTTGGACAGTCCAGATGAAATGACATCAATGTGGAT 119977

      97 AspIleGluThrProAspGluThrAspSerLeuGluPheLeuGlyAsnGl 113
         ::::::::::|||||||||||||::|||||:::'|||:::   |||:::||
  119978 GAACTTGATACCCCCGATGAAGCAGATTC:|||:AGIACACTGGCCATGG 120027

     113 yAsnGluLeuGluTrpGlu 119
         |:::::::|:::|||:::
  120028 TAAGTCAC1AAG1TGCCAA 120046

seq_name: gb_htg5:AC014963

seq_documentation_block:
LOCUS       AC014963   126121 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014963
VERSION     AC014963.1  GI:6436372
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 126121)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211484 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..126121
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    35005 a  26506 c  26872 g  36738 t
ORIGIN


alignment_scores:
            Quality:  236.00              Length:    340

              Ratio:  1.380                Gaps:     10
 Percent Similarity:  50.294   Percent Identity: 22.941

alignment_block:
 US-09-174-937-2 x AC014963   ..

 Align seg 1/1  to: AC014963  from: 1  to: 126121

      66 ProGluIleAsnIleSerLeuAspGlnSerGluGlySerLeuLeuSerAs 82
         |||   ||||||      :::|||:::|||
  16777 CCCGCCATCAATCCGATCGTCGATAACTCG          ..........  16806

      82 pAspPheLeuAspThrProAspAspLeuAspIleAsnValAspAspIleG 99
                |||   |||:::              ::::::   ||||:::|
  16807 ..........GATGAGCCCCAACCCAGICICICCGAICTGCACGACIICG 16856

      99 luThrProAspGluIhrAspSerLeuGluPheLeuGlyAsnGlyAsnGlu 115
         ||         |||   |||:::   |:::|||:::
  16857 AGCCGAAGCTAGAGTTTGATGATACCGAGCTGCTGGCC............  16904

     116 LeuGluTrpGluAspAspThrProValAlaThrAlaLysAsnMet..... 130
                    :::|||:::  :::     |||:::
  16905               CCATCTCCGCTTGAAAGTAAATTGAATATTTAAAG 16939

     131 ....                ....                ProGlyAspS 134
                                                 |||   |||
  16940 AATCTACAAAATCATAGTCAAACIAAICTCIICGAITTTTCCCAGAAGAIG 16989

     134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 150
            :::   ||||||   :::      :::|||:::
  16990 TI......ATGGTCGGTGATTTCGTTCTGGCCGAAGGTAAG.........  17034

     151 GlyArgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAs 167
                |||||.:.....      ...|'|    '|:::
  17035 .........TGGCGGAATCTACTTCTCCIGGCCAATGAATAACGTTTCGC 17085

     167 pLeuHisMet.........IleArgProTyrMetLysVal.......... 177
          |||  ...  .      ..:||'     '|.
  17086 ATTACGCCTAGATCCAGAACTTGAGCCCGAGGAAGATGTAAATCCGCTGG 17135

     177                ..   ..   ..                ..... 177

  17136 AGGACGACTTCGAGGACCAACTCAGGGAGCAGTCTGAAAACTTTCAGACT 17185

     178             ValThrHisGlyGlyTyrTyrGlyGluGlyLeuAsnAl 194
                        |||          :::|:|   :::
  17186 CCCAGAAACAAATGCGATTTCCTGGGAACCGATAAGCAGGGICGCCACAI 17235

     190 aIleIleValPheAlaAlaCysPheLeuProAspSerSerLeuProAspT 207
         .   :::     |||        |||.....|      :
  17236 CTTTGGCATTTATGCATCCCGCTIICCCGAAAAATCCCAACTCGAAGGAI 17285

     207 yr.....................HisTyrIleMetGluAsnLeuPhe 215
         ::                    :::::::::::::::::||:::
  17286 TTGTGCGGTAGGATTCATATTCGAGTACATTTTTGTTCAATCATTAATC 17335

     216 LeuTyr.................ValIleSerSerLeuGluLeuLeuVa 226
            :::                 :::|||::::::::|||   :::''
  17336 TATTTCCATTTGTCTATCAGGGAAATCATCAAGAGATCGAACCTTTTGI 17385

     226 lAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgA 243
         |   :|||||  :   ||||||          ::      :::   :
  17386 GGAGAACGACTACATTTTGGTTTACTTCCATCAAGGCCTCAAGGAGGACA 17435

     243 rgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArg 259
         ::   |||:::   :  |||  :::  |.|::   :::''''.
  17396 ACAAACCATCCGCACAGTTCCTGTGGAATTCGTACAAGGAGCTCGATCGC 17445

     260 ArgLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpPh 276
         ::::::|||||||||||||||::|||   :::|||||||||:::|||'
  17446 AACTTCCGCAAGAATCTGAAGACGTTGTATGTGGTTCATCCGACGTGGTT 17495

     276 eIleArgThrValLeuAlaIleSerArgProPheIleSerValLysPhel 293
```

```
                                                                     ||||||
 17496 CATACGGGTCATTTGGAACTTCTTTAGCCTGTCATTAGCGACAAGTTCC 17545

   293 leAsnLysIleGlnTyrValHisSerLeuGluAspLeuGlnLeuIle 309
         :::|||:::   |||:::   |||||||:::::|||:::|||   :::
 17546 GCAAGAAGCTAGTCTACATCTCCAGCTTGGACAAGTTGCGCCAGGCGCTC 17595

   310 ProMetGluHisValGlnIleProAspCysValLeuGlnTyrGluGluGl 326
       ::::::  .........||||||  ...  ...  .....
 17596 GGCCTGAACAAGCTGAAGTTGCCGGACAACATCTGCGACCTG...GACGA 17642

   326 uArgLeuLysAlaArgArgGluSerAlaArgProGlnProGluPheValL 343
       :::::|:::   |||:::  :::  |  |||:::   :
 17643 CAAACTGAACCCCAGCCGGAAGCCATCGACACTGCGCCCCAGCAGCAATA 17692

   343 euProArgSerGluGluLys 349
       ::    ||||||||||||
 17693 TAAATGCATCCCCGCAGCAG 17712

seq_name: gb_pri2:HSU62794

seq_documentation_block:
LOCUS       HSU62794     1378 bp    mRNA            PRI       13-AUG-1997
DEFINITION  Human CDC42 GAP-related protein mRNA, partial cds.
ACCESSION   U62794
VERSION     U62794.1  GI:2326170
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Zheng,Y., Kuang,W.J., Hart,M.J., Evans,T.,
            Cerione,R.A. and Ashkenazi,A.
  TITLE     Cloning and expression of a human CDC42 GTPase-activating protein
            reveals a functional SH3-binding domain
  JOURNAL   J. Biol. Chem. 269 (35), 22031-22034 (1994)
  MEDLINE   94075275
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Lancaster,C.A., Taylor-Harris,P.M., Self,A.J., Brill,S., van
            Erp,H.E. and Hall,A.
  TITLE     Characterization of rhoGAP. A GTPase-activating protein for
            rho-related small GTPases
  JOURNAL   J. Biol. Chem. 269 (2), 1137-1142 (1994)
  MEDLINE   94117418
REFERENCE   3  (bases 1 to 1378)
  AUTHORS   Barfod,E.T. and Ashkenazi,A.
  TITLE     An effector function for Cdc42GAP that modulates the mitotic
            spindle
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 1378)
  AUTHORS   Barfod,E.T., Chan,M., Kuang,W-J. and Ashkenazi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-1996) Molecular Oncology, Genentech, 460 Point
            San Bruno Blvd., S. San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="CMK"
                     /cell_type="megakaryocyte"
     CDS             <1. .672
                     /note="CDC42 GAP-related protein encoded by GenBank Accession Number U02570;
                     insertion in the rhoGAP active site renders active site
                     boxes 2 and 3 out of frame and untranslated"
                     /codon_start=1
                     /product="CDC42 GAP-related protein"
                     /protein_id="AAB66594.1"
                     /db_xref="GI:2326171"
                     /translation="GCNELPSHSSLGEGDSVSKKKSMHWDPSVAPVPPRLLLSPFFCL
                     CRYKVNIKALYIVHPTMPTFT LILFKPLISFKPGQKIFYVNYLSELSEHVKLEQLGI
                     PRQVLKYTQFLKSIQKSPAIAKPMPPPPPPLPNQQFGVSLQHLQEKNPEQEPIPIVLP
                     ETVAYLQAHALITERIPRPSANTQNVREV__KYNMI CCPFSGCAKILSPCPHPRAACG
                     FRPVQ"
     misc_feature    370. .402
                     /note="encodes SH3-binding domain capable of binding src
                     and p85"
     misc_feature    465. .543
                     /note="active site box 1"
     misc_feature    585. .640
                     /note="55 nucleotide insertion into sequence identical
                     with CDC42GAP; insertion between active site boxes 1 and 2
                     shifts translational frame 4, 1 that a stop codon occurs
                     at 670, and active site boxes 2 and 3 are out of frame"
     misc_feature    671. .763
                     /note="out-of-frame active site box 2"
     misc_feature    821. .949
                     /note="out of frame active site box 3"
     polyA_site      1378
                     /note="21 A nucleotides"
BASE COUNT      288 a    468 c    328 g    294 t
ORIGIN


alignment_scores:
           Quality:  173.00            Length:    111
             Ratio:  2.437               Gaps:      2
 Percent Similarity:  68.964   Percent Identity: 56.034

alignment_block:
 US-09-174-937-2 x HSU62794   ..

 Align seq 1/1  to: HSU62794  from: 1  to: 1378

     260 ArgLeuArgLysAsnLeuLysSerL...IleIle'alHisPr.S.rIrpIh 278
         :  ::|||:::: ::: : :::                  :::   :
     136 AGGTACAACAAGAACATCAACGCTTTCTACATCTTCTATTCAATCAATCTT 185

     279 eIleArgThrValLeuAlaIle.:rsProPhIl-SerValLysPhe 295
         :::  |::: ::  :::  ::: :::  :  :::  :  :
     186 CATCAAAACTCTGCTCATCCTCTTCAATTTCCTTATTAGCTTAAGTTTT 235

     293 leAsnLysIleGlnTyrValHisSerLeuGluAspLeuGlnLeuIle 309
         :::  :  :::  :  :::  ::::::  :::  :  :::
     236 GGCAGAAGATCTTCTATGTGAATTAGTGAGCTAGTGAGTAATAGTAGTTT 285

     310 ProMetGluHisValGlnIleProAspCysValLeuGlnTyrGluGluGl 326
         ::: ||::::::  :::  :  :::  ::::  :::  :
     286 AAGCTCGAGCAGCTGGGCATCCCTGCCGAACTCTTCTCAAATAT...GATGA 332

     326 uArgLeuLysAlaArgArgGluSerAlaArgProGlnProGluPheValL 343
         :  ::::  ::: ::  :  :::  :  :::  :  :
     333 CTTCCTGAAATCCACACAGAAGAGTCCCTCCCACAGCTCCCAATCTTATTT 382

     343 euProArg.....                                      349

     383 CCCCAGGGCCCCCCCTGCCCAACTAGTAG..TTGAG.CTTTTTTAGTAT 432

     346 ...SerGluGluLysProGluValAlaProVal 357
         ...:::.::::::::::  :  :::
     433 CTCCAGGAGAAGAATCCAGAGCAGGAGCCCATT 467

seq_name: gb_htg1:CEY62F5

seq_documentation_block:
LOCUS       CEY62F5    210566 bp    DNA             HTG       ... 11-1999
DEFINITION  Caenorhabditis elegans chromosome II clone Y62F5  *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL023633
VERSION     AL023633.2  GI:5102738
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
```

```
REFERENCE   1  (bases 1 to 310565)
  AUTHORS   Williams,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 22, 1999 this sequence version replaced gi:4938509.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..310565
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y62F5"
BASE COUNT    94989 a  58692 c  57839 g  94851 t   4194 others
ORIGIN


alignment_scores:
            Quality:  179.00              Length:    321
              Ratio:  1.104                 Gaps:     14
 Percent Similarity:  47.975   Percent Identity: 23.988

alignment_block:
 US-09-174-937-2 x CEY62F5

 Align seq 1/1  to  CEY62F5  from  1  to  310565

      43 AspThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLy 59
         ||||||||:  :|||||||:      :::  :::
   72807 GACACGAAACATCCTCCCGATCCATCAGGTCCATCAGAT      ..... 72845

      59 sArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerG 76
                        |||:::|||| |:::::     ::::::
   72846                GCCAACATAAAICCTGGAGAATTCGAAGATC 72876

      76 luGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuAsp 92
            :::      :::|'|||||  :||    '|||||||||||||:::
   72877 CGGTGGATGGCTATGCCGATGATGACCTTCTCACACCGGATGACCTAAGC 72926

      93 .IleAsn...ValAspAspIleGluThrProAspGluThrAspSerLeuG 108
         ::::::   ...    |||...          ...          :
   72927 CGTAAGTGGAATAATAATTATTGATATTTAATTTAATAGAAAACTGGTTT 72976

     108 luPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProVal 124
          |||          ::::::          ::: :::      :
   72977 ATTTC            AGTAAGTGTCACTGGAATGACGCCATCTCGATC 73014

     125 AlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuPheGlyAspGl 141
         ::::::                                  |||:::
   73015 GTTTCT              . .                 GGAAACTG 73028

     141 yThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrValI 158
               :::|||          :     ||||||
   73029 AAGAATCGAGAC              .   AAGCTTGGAGGTGTTGAA 73060

     158 leIleGlyGluGlnGluHisArg.IleAspLeuHisMetIleArgProTy 174
           :::  ::::::  :::|||| ||||::   ||   |||
   73061 CGTTCGAAGAICTTTTCAACGACATAICAGCTCACGAAATT . .     73102

     174 rMetLysValValThrHisGlyGlyTyrTyrGlyGlu.     GlyLeuA 189
         :::::::||:::::::  |||         | ::::.      ...|||

   73103 .ATTCAAGTAATTGCTGATGGGGATCGAGTTGGAAGGTAAATTAATTTGA 73151

     189 snAlaIleIleValPheAlaAlaCysPhe..................... 198
         ||      |||   :::     ||||||
   73152 ATCCAAAAATTAATTATCTTTTTTGTTTTAATGAATCATAGTTTGCAGGC 73201

     199 . . . . . ..LeuProAspSerSer                    213
                     |||      |||
   73202 CAATTGTTGTCGTATATGCTTATCGGCTCCCATCAAGCAAGGAAATAGAT 73251

     204 .LeuProAspTyrHisTyrIle........................... 218
         ::|||||  . .  |||
   73252 CATGCCAGACTTCTACAGTATTTGGTACAAATTATTGATAAGGTTGGGAA 73301

     211 ........MetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeu 224
                 |||.........   :::       ::::::
   73302 TTTTTAATAAATATAATATTTTGATAAAAATAAAATTTGCAATTTTCCAG 73351

     225 LeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProAr 241
         :::|||   :::||||||   ||||| | |:::::   :::
   73352 ATCGTTGACCAAGACTACACAATTGTCTATTTCACTATGGCCTACGAAG 73401

     241 gArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleA 258
         ::::::   |||   :::   |||| |  ::::::|||::::::::::
   73402 TCACAACAAACCACCGGTACGGTGGCTGTTCCAGGCTTACAAACAACTTG 73451

     258 spArgArg. Leu. ....... ................. 261
         |||||||
   73452 ATCGTAGGTAAGTATTTGAATTCAATTCATCTCTTCTCTTTCAAAAAAAC 73501

     262 .............ArgLysAsnLeuLysSerLeuIleIleValHisPro 278
                     :::||||||||||| ||:::|||   :::||||||||
   73502 CGGTTTCCAGATTCAAAAAGAACCTGAAAGCGCTCTACGTAGTGCATCCA 73551

     274 SerTrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerVa 293
         :::   ||||||||   ..........   :::   ||||||||
   73552 ACACGATTCATCCGCATTATCTTCTCGCTGTTCAAAGGTTTCATATCGAG 73601

     290 lLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeuGluG 307
         |:::::   |:|||:.....|'|||'     ...:::::|||'|::
   73602 CAAGTTTGAGAACAAGTTCCACTATGTCATGTGTATAGACGAGCTGGAGA 73651

     307 lnLeuIle 309
         .:   :::
   73652 ACGCACTT 73659

seq_name: gb_in1:CEW02B12

seq_documentation_block:
LOCUS       CEW02B12    34305 bp    DNA             INV       10-NOV-1999
DEFINITION  Caenorhabditis elegans cosmid W02B12, complete sequence.
ACCESSION   Z66521
VERSION     Z66521.1  GI:1469136
KEYWORDS    HTG; Alanyl-tRNA synthetase; Esterase; Kinensin like protein;
            Mitochondrial RNA splicing like protein, N-acetyllactosamine
            synthase; Phospholipase; Pre-mRNA splicing factor like protein;
            Serine/threonine protein kinase.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 34305)
  AUTHORS   Wilson,P., Ainscough,P., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
```

```
   TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
             elegans
   JOURNAL   Nature 368 (6466), 32-38 (1994)
   MEDLINE   94150718
REFERENCE    2  (bases 1 to 34305)
   AUTHORS   Swinburne,J. and Ainscough,R.
   TITLE     Direct Submission
   JOURNAL   Submitted (27-OCT-1995) Louis, MO 63110, USA. E-mail:
             jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT      On Jul 29, 1996 this sequence version replaced gi:1044853.
             Coding sequences below are predicted from computer analysis, using
             predictions from Genefinder (P. Green, U. Washington), and other
             available information.
             For a graphical representation of this sequence and its analysis
             see:-
             http://webace.sanger.ac.uk/cgi-
             bin/display?db=wormpep&class=Sequence&object=W02B12
             Current sequence finishing criteria for the C. elegans genome
             sequencing consortium are that all bases are either sequenced
             unambiguously on both strands, or on a single strand with both a
             dye primer and dye terminator reaction, from distinct subclones.
             Exceptions are indicated by an explicit note.
             IMPORTANT:  This sequence is NOT necessarily the entire insert of
             the specified clone.  It may be shorter because we only sequence
             overlapping sections once, or longer because we arrange for a small
             overlap between neighbouring submissions.
             IMPORTANT: This sequence is not the entire insert of clone W02B12.
             It may be shorter because we only sequence overlapping sections
             once, or longer because we arrange for a small overlap between
             neighbouring submissions.
             The true right end of clone W02B12 is at 4181 in
             sequence Z66519.
             The true left end of clone B0334 is at 34205 in this sequence. The
             start of this sequence (1. .104) overlaps with the end of sequence
             AL031269.
             The end of this sequence (34272   34305) overlaps with the start of
             sequence Z66519.
FEATURES             Location/Qualifiers
     source          1   34305
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="W02B12"
     gene            4120. .6453
                     /gene="W02B12.1"
     CDS             join(4120. .4132,4598. .4707,4755. .4872,4923. .5089,
                     5135. .5301,5556. .5738,5788. .5860,6000. .6104,6154. .6261,
                     6313. .6453)
                     /gene="W02B12.1"
                     /note="similar to phospholipase, cDNA EST yk470h3.3 comes
                     from this gene; cDNA EST yk470h3.5 comes from this gene;
                     cDNA EST yk539g12.3 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA91893.1"
                     /db_xref="GI:3880427"
                     /db_xref="SPTREMBL:Q09119"
                     /translation="MRWLVTTLLCCSLAHGLKDLGAPNWSCDASVMAKSKKVPTSAHS
                     VRFADIKVIGALSDSLIAANGAGAQPGDPLAVILQYRGLAFQIGGDKSLDEHITVANV
                     LRKFNPTLVGASKGIGSFNVWDVSHLNMGVPGAESKDIIGQAPALVNTMHAHSEIVVK
                     EDWKLVNIFIGANDICVYCFTPYFNSTALHGNATFEKNIIAAVQILQDNLPPRTNFKF
                     KLPSRISVCKTFSSWPFFKNSNIYSPTIVSLTGMFNMPMLPKIDKKKYPCEGLHTFEC
                     DCESNKQFTDDDIQGVCFSYWNAEKDIQNIGLFDNKDDFIFVVQPFFNGILDPPYASP
                     GVVDMIFFAPDCFHFSAYGHSNIGMHLWNTIVQPVGFKQISVNLSDPSVGLHCPSINC
                     PFFPTIKNSNNCASHFTLSELD"
     gene            complement(6752   8004)
                     /gene="W02B12.2"
     CDS             complement(join(6752. .6845,6894   7190,7239   7315,
                     7605. .7862,7915. .8134))
                     /gene="W02B12.2"
                     /note="similar to pre-mRNA splicing factor like protein;
                     cDNA EST EMBL:T01363 comes from this gene; cDNA EST
                     EMBL:D65797 comes from this gene; cDNA EST EMBL:D72951
                     comes from this gene; cDNA EST EMBL:D63365 comes from this
                     gene; cDNA EST EMBL:D74120 comes from this gene; cDNA EST
                     EMBL:D75914 comes from this gene; cDNA EST EMBL:D74773
                     comes from this gene; cDNA EST EMBL:D74770 comes from this
                     gene; cDNA EST EMBL:D74114 comes from this gene; cDNA EST
                     EMBL:D74897 comes from this gene; cDNA EST EMBL:D77896
                     comes from this gene; cDNA EST EMBL:D74171 comes from this
                     gene; cDNA EST EMBL:D77345 comes from this gene; cDNA EST
                     EMBL:D58083 comes from this gene; cDNA EST yk289c5.3 comes
                     from this gene; cDNA EST yk334a4.3 comes from this gene;
                     cDNA EST yk334a4.5 comes from this gene; cDNA EST yk745f7.  comes
                     from this gene"
                     /codon_start=1
                     /protein_id="CAA91894.1"
                     /db_xref="GI:3880428"
                     /db_xref="SPTREMBL:Q22121"
                     /translation="MTSLIIVLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
                     DRDADDAVHDLNSKELSSRPVILEPPPPPKYTNSERSSSPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     DLRDAMNKLQGEILNGRKLKIDETRLRSSSPSPPPPSPPSPSPPPPSPPPPARRRSPGS
                     LRSLPPPPSASAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     GSASP"
gene             9070. .10722
                 /gene="W02B12.3"
CDS              join(9070. .9192,9246. .9478,9553. .9612,10114. .10187,
                 10240. .10519,10579. .10722)
                 /gene="W02B12.3"
                 /note="similar to pre-mRNA splicing factor like protein;
                 cDNA EST EMBL:D71552 comes from this gene; cDNA EST
                 EMBL:D73263 comes from this gene; cDNA EST EMBL:D74207
                 comes from this gene; cDNA EST EMBL:D7725   comes from this
                 gene; cDNA EST EMBL:D 8077 comes from this gene; cDNA EST
                 EMBL:D10112 comes from this gene; cDNA EST EMBL:D 7321
                 comes from this gene; cDNA EST EMBL:D 8072 comes from this
                 gene; cDNA EST yk479g12.3 comes from this gene; cDNA EST
                 yk479g12.5 comes from this gene; cDNA EST yk454f4.3 comes
                 from this gene; cDNA EST yk454f4.5 comes from this gene;
                 cDNA EST yk454f4.3 comes from this gene; cDNA EST
                 yk454f4.5 comes from this gene; cDNA EST yk44   comes
                 from this gene; cDNA EST yk44   comes from this gene;
                 cDNA EST yk401b2.3 comes from this gene; cDNA EST
                 yk401b2.5 comes from this gene; cDNA EST yk   g11.3 comes
                 from this gene; cDNA EST yk   g11.5 comes from this gene;
                 cDNA EST yk   a   comes from this gene; cDNA EST
                 yk303a6.5 comes from this gene; cDNA EST yk   g12.  comes
                 from this gene; cDNA EST yk   g12.  comes from this gene;
                 cDNA EST yk282b11.3 comes from this gene; cDNA EST
                 yk282b11.5 comes from this gene; cDNA EST yk   g   comes
                 from this gene; cDNA EST yk   g   comes from this gene;
                 cDNA EST yk229b9.3 comes from this gene; cDNA EST yk   g12.  comes
                 from this gene; cDNA EST yk   g   comes from this gene;
                 cDNA EST yk570b9.3 comes from this gene; cDNA EST yk   g   comes
                 from this gene; cDNA EST yk   g   comes from this gene;
                 cDNA EST yk   g   comes from this gene; cDNA EST
                 yk595a1.3 comes from this gene; cDNA EST yk   g   comes
                 from this gene; cDNA EST yk   g   comes from this gene;
                 cDNA EST yk   g   comes from this gene; cDNA EST
                 yk637g12.3 comes from this gene; cDNA EST yk   g   comes
                 from this gene; cDNA EST yk   g   comes from this gene;
                 cDNA EST yk   g   comes from this gene; cDNA EST
                 yk675e1.3 comes from this gene; cDNA EST yk   g   comes
                 from this gene"
                 /codon_start=1
                 /protein_id="CAA91895.1"
                 /db_xref="GI:3880429"
                 /db_xref="SPTREMBL:Q22121"
                 /translation="MTASIIVLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
                 DKRDAEDAVHDLNSKELSSRPVILLYSKFFSSSPSPSSSPSPSPSPSPSPSPSPSPSP
                 GRDRFDRYDRGPPPPSSPPSPPSPPPPYYYENLSPPSPPLPPLLVFSLGVEPIYAE
                 AHKRPNDALLSVAIISLKRIIRKDDTWCINGSPPIKNITLPLAIILGPNSRSRSRS
                 RSRDRRRSRSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
```

```
                VDRSPSPPPGSRSPSEKGSPPSPSASPMDNGDGDN"
     gene       complement(11094. .14445)
                /gene="W02B12.4"
     CDS        complement(join(11094. .11199,11252. .11517,11739. .11804,
                11859. .12049,12121. .12337,12380. .12510,12572. .13073,
                13651. .13704,13748. .13894,14236. .14445))
                /gene="W02B12.4"
                /note="similar to esterase"
                /codon_start=1
                /protein_id="CAA91397.1"
                /db_xref="GI:3880431"
                /db_xref="SPTREMBL:Q23123"
                /translation="MNWVQTPLLIFLTTLVFPCTADPLETSADSPVLSTGLGYIPGP
                PVSTKSGFLVDYFTGVPPAEKPEPLKKSKLVNPWVDIPNTQQYPPKCIPAAPPVNISE
                FSEDCLYMNILKPSGLNIFRKYPVVVFVHGGGFQEVLFFSILNVLGDGNDASCDAFAE
                HFVSKEMIVVTIQYRLGMLGFLSLGCSDPLIPCNLGLWDMVTAFPFINATIENFNGDI
                NNMILMGHSAGAMAVSLHSMSPISSVHFKKVIQLSSSSWGLIFYKDQNIVQIKKILKD
                LNCDISSAKSIVDCVNSTSLDSMYSAQVCAHPFSFEYIFSPVETGYITFAPKIANFRI
                CVIPLFVQEILKTRVDTTLWPEFGDELIPDIPDNMVKNTQNQKVLMGVMILESLYFSE
                FRFFVNCQIYKKNVGFVNELKVCTSYFRAYLKSSPAIFPYIHRDTVEEYINSTIAIQY
                GIGSESAYDMTSAYYLSANATDNDYGYFMWQPSKVSARQMFLKSTVTLVALEFPKKID
                SNTGFDIPIIREAPARPSQVSPDLYLYHFPYFNPAEFDVDFPVRATYHCHEFPYIWGVY
                KDNYFEFNDDDRTVSAFLITSLINFIKTGNPSSSEFNWPLYDSSVSHIVIGPKPTTDK
                RLLTL"
     gene       complement(15419. .19110)
                /gene="W02B12.6"
     CDS        complement(join(15419. .15490,15540. .15723,16310. .16580,
                16632   16854.16992   17155.17204   17288,17446   17868,
                17926   18642.18646   18986,19033   19110))
                /gene="W02B12.6"

alignment_scores:
            Quality:  170.00             Length:    321
              Ratio:  1.104                Gaps:     14
 Percent Similarity:  47.975   Percent Identity: 23.988

alignment_block:
 US-09-174-937-2 x CEW02B12

 Align seq 1/1  to: CEW02B12  from: 1  to: 34305

      43 AspThrSerSerProProAsnThrLeuAsnPheAsnGlyAlaHisArgLy 59
         ||||||||:   |||||||:::     :::   :::
   19989 GACACGAAACAICCICCGGAICCAICAGGICCAICAGAT.......... 20027

      59 sArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSerG 76
                        |||:::|||.||:::::::     ::::::
   20028 ..................CCCAACATAAATCTTGGAGAATTCGAAGATC 20058

      76 luGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuAsp 92
            :::     :::|||||||   |||   .|||||||||||||:::
   20059 CGGICGAICGCCIAICCCGAICGATGACCTTCICGACACCGGATGACCTAAGC 20108

      93 .IleAsn...ValAspAspIleGluThrProAspGluThrAspSerLeuG 108
         ::::::   :::      |||:::         :::         :::
   20109 GGIAAGTGGAATAATAATTATTGATATTTAATTIAATAGAAAACTGGTTT 20158

     108 luPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProVal 124
           |||         ::::::   :::.|||:::|||         :::
   20159 ATTTC       .     AGTAAGTGTCACTCGAATGAGCCATCTCGATC 20196

     127 AlaThrAlaLysAsnMetProGlyAspSerSerAspLeuPheGlyAspGl 141
         ::::::                                  |||:::
   20197 CTTTAT                                    CCAAACTG 20210

     141 yThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrValI 158
               :::|||                 |     ||||||
   20211 AAGAATTCGAGAC     .       ...   AGAGCTTGGAGCTGTTGAA 20242

     159 leIleGlyGluGlnGluHisArgIleAspLeuHisMetIleArgProTy 174
         :::   ::::::   :::|||.|||:::    ||   |||
   20243 CCGTTCGAAGATCTTTTCAACGACATATCAGTTCACGAAATT........ 20284

     174 rMetLysValValThrHisGlyGlyTyrTyrGlyGlu.....GlyLeuA 194
         .:::::|||:::::.   |||         |||:::      ::::|||.
   20285 .ATTCAAGTAATTGCTGATGGGGATCGAGTTGGAAGGTAAATTAATTTGA 20333

     189 snAlaIleIleValPheAlaAlaCysPhe..................... 198
         ||      |||   :::       ||||||
   20334 ATCCAAAAATTAATTATCTTTTTTGTTTTAATGAATCATAGTTTGCAGCC 20383

     199 .................LeuProAspSerSer................... 209
                          |||       |||
   20384 CAATTGTTGTCGTATATGCTTATCGGTCGCATCAACCAAGCAAATAGAT 20433

     204  LeuProAspTyrHisTyrIle............................ 217
          ...||||||:::::   |||
   20434 CATGCCAGACTTCTACAGTATTTGGTACAAATTATTGATAAGGTTGGGAA 20483

     211        . MetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeu 224
                     |||:::::::   :::       ::::::
   20484 TTTTAATAAATATAATATTTTGATAAAAATAAAATTTGCAATTTTCCAG 20533

     225 LeuValAlaGluAspTyrMetIleValTyrLeuAspGlyAlaThrProAr 241
         :::|||  ::::|||||||  |||||.| ::::::  :::
   20534 ATCGTTGACCAAGACTACACAATTGTCTATTTTCACTATGGCCTACGAAG 20583

     241 gArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleA 258
         .:::::   |||   :::   |||'|:   :::::::||::::::::::
   20584 TCACAACAAACCACCGGTACGGTGGCTGTTCCAGGCTTACAAACAACTTG 20633

     258 spArgArg  Leu.                          ..............  261
         ||||||||
   20634 ATCGTAGGTAAGTATTTGAATTCAATTCATCTCTTCTCTTTCAAAAAAAC 20683

     262 ..............ArgLysAsnLeuLysSerLeuIleIleValHisPro 273
                        :::|||||||||||||:::|||   :::|||:'|||
   20684 CGGTTTCCAGATTCAAAAAGAACCTCAAAGTGCTCTACGTATTTTATCTA 20733

     274 SerTrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerVa 290
         :::  |||||||||  ..............  :::  |||||||||
   20734 ACACGATTCATCCGCATTATCTTCTCGCTGTTCAAACGTTTCATATCGAG 20783

     290 lLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeuGluG 307
         ||||||  ||||||:::::::|||||:::      ::::::::|:::|:
   20784 CAAGTTTGAGAACAAGTTCCACTATGTCATGTGTATAGACGAGCTGGAGA 20833

     307 lnLeuIle 309
         ::   :::
   20834 ACGCACTT 20841

seq_name: gb_in2:AF091997

seq_documentation_block:
LOCUS       AF091997     2234 bp    DNA             INV       11-FEB-1999
DEFINITION  Plasmodium falciparum rhoGAP protein gene, partial cds.
ACCESSION   AF091997
VERSION     AF091997.1  GI:6018182
KEYWORDS    .
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 2234)
  AUTHORS   Soubes,S.C., Liu,X. and Miller,L.H.
  TITLE     Representational diff[illegible] analysis of cDNA between two clones
            of Plasmodium falciparum
  JOURNAL   Mol. Biochem. Parasitol. 101 (1-2), 217-231 (1999)
  MEDLINE   99339401
REFERENCE   2  (bases 1 to 2234)
  AUTHORS   Soubes,S.C., Liu,X. and Miller,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-1998) Lab. of Parasitic Diseases, NIAID,NIH, 4
            Center Drive, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1..2234
                     /organism="Plasmodium falciparum"
```

```
                     /db_xref="taxon:5833"
                     /clone="A4"
                     /country="Brazil"
     mRNA            join(<264..1111,1263..>1877)
                     /product="rhoGAP protein"
     CDS             join(<264..1111,1263..>1877)
                     /codon_start=3
                     /product="rhoGAP protein"
                     /protein_id="AAF01782.1"
                     /db_xref="GI:6018183"
                     /translation="YDTLPPKYKKNLKNLYLVHSGFLSKTLLTIVTPFISPKFWKKVE
                     YIEKLEDLFLKLNIKASKYLKYFPYIVQRNEEVMLGGQPISPFGADLEILCQRFGKKY
                     LHFKHIPSILIDFLTHLKKPEIINTKDLFNLQADAATLYGIVGDIEYGEPTTDFNNIP
                     SLVCSFKLFLNTQKHGLLGKDAFSRLHYLKSTSASDKVIKLNLSKLYEKLSQGIKQCV
                     LCLLHFFKTISEHASENNMTILALSKIFAPTFFRSKTPNAIFTECISLANKCMQMIIQ
                     NPDILMNPEKYNEQSSESISSYESGSENNLTEDEDKSDESKSDESKSDESKSDDSKSD
                     ESKSDESKSDESKSDESKSDESKSDESKSDDSKSEECKDENNKINNESDEDSNESYSS
                     KESSSKPSSLDQTSEYSSSILSDKVIKKKVSIKKSTSKRLSFDHSIKSEKSGDSKNSF
                     LSEKTDEQDDSYEDSEKEEEGKKSPKENLKKEPSSDN"
BASE COUNT      939 a    239 c    271 g    785 t
ORIGIN


alignment_scores:
           Quality:  143.50              Length:     99
             Ratio:  2.315                 Gaps:      1
 Percent Similarity:  62.626   Percent Identity: 29.293

alignment_block:
 US-09-174-937-2 x AF091997   ..

 Align seg 141  to: AF091997  from: 1  to: 2234

     207 TyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLeuGl 223
         :||:   ||||||          :::.....:||:::::       |||
     176 TATATATATATATATATATATATATATATATATATATTTATTTATTTATT 225

     223 uLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrP 240
          :::::::::          :::::::::.....:::::
     226 TATTATTTTATTTTATTTTTTTTTTGTTTTATTTGTAGGG.......... 265

     240 roArgArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMet 256
                                                 |||:::
     266 ....    .    .  .  .. .... ............TATGATACA 274

     257 IleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHisPr 273
         :::   |||:::   :::||||||||||||||:::|||   :::||||||
     275 TTACCCAGGAAATACAAAAAGAAIITAAAAAACTTATATCTGGTTCATAG 324

     273 oSerTrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerV 290
          :::::::::   ...|||...:|||:::||:      ||||||||||||
     325 TGGTTTTTTATCAAAAACTCTTTTGACCATTGTAACTCCATTTATATCCC 374

     290 alLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeu 305
          |:|||   ...|||.. ..|||::::::::::|||||||||||||
     375 CAAAATTTTGGAAGAAGGTGGAATATATAGAAAAGTTAGAAGATTTA 421
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2000, 17:41.30 , Search time 3770.17 Seconds
(without alignments)
3080.482 Million cell updates/sec

Title: US-09-174-937-1
Perfect score: 3076
Sequence: 1 ccacgcgtccgcccacgcgt.....ggcatggtggtgcatgcctg 3076

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	437.4	14.2	495	47	AI498192	AI498192 tm90g12.x
	2	393.8	12.5	437	20	T08680	T08680 EST06572 In
c	3	393.2	12.5	461	47	AI480208	AI480208 tm72f02.x
	4	349.8	11.4	368	20	T05645	T05645 EST03534 Fe
	5	331.6	10.8	384	20	M78376	M78376 EST00524 Fe
	6	321	10.4	877	82	AQ739838	AQ739838 HS_5505_A
	7	309.8	10.1	335	20	T09124	T09124 EST07017 In

	8	307.6	10.0	590	104	AQ586588 AQ586588 CITBI-E1-
	9	305	9.9	449	100	AQ312765 AQ312765 RPCI11-1M
	10	300.8	9.8	511	99	AQ215135 AQ215135 HS_3249_B
c	11	299.2	9.7	660	102	AQ381318 AQ381318 RPCI11-13
	12	298	9.7	922	87	AQ748222 AQ748222 HS_5536_A
	13	296.6	9.6	795	102	AQ346434 AQ346434 RPCI11-10
c	14	296	9.6	558	81	B68229 B68229 CIT978SK-A-
c	15	292	9.5	440	95	AA601278 AA601278 no15b09.s
	16	291.4	9.5	534	104	AQ543763 AQ543763 RPCI-11-3
c	17	289.6	9.4	555	104	AQ541476 AQ541476 RPCI-11-3
	18	289	9.4	626	98	AQ878847 AQ878847 HS_3144_B
	19	288.4	9.4	590	87	AQ749543 AQ749543 HS_5575_A
c	20	287.8	9.4	761	92	AQ742286 AQ742286 HS_5383_B
	21	287.4	9.3	575	104	AQ583736 AQ583736 RPCI-11-4
	22	283	9.2	710	92	AI110760 AI110760 HA0203 Hu
	23	283	9.2	710	74	AF074667 AF074667
c	24	282	9.2	529	87	AA680243 AA680243 ac86a01.s
c	25	281.4	9.1	560	61	AL042539 AL042539 DKFZp434I
	26	281.4	9.1	622	104	AQ541531 AQ541531 RPCI-11-3
c	27	281.2	9.1	431	61	AI816537 AI816537 au48g06.x
c	28	281.2	9.1	5528	79	AB012089 AB012089 Homo sapi
	29	280.4	9.1	613	102	AQ419352 AQ419352 RPCI-11-2
	30	279.8	9.1	942	87	AQ743787 AQ743787 HS_5501_A
c	31	279.6	9.1	813	89	AQ898730 AQ898730 HS_3129_B
	32	279.6	9.1	672	102	AQ394009 AQ394009 CITBI-E1-
	33	279	9.1	491	60	AI791227 AI791227 oe21g10.y
	34	279	9.1	489	87	AQ797930 AQ797930 HS_2084_A
c	35	278.6	9.0	914	90	AF102051 AF102051 Homo sapi
c	36	278.6	9.0	475	81	B63124 B63124 CIT978SK-A-
	37	278.2	9.0	676	61	AL040847 AL040847 DKFZp434G
	38	278.2	9.0	553	90	AQ037614 AQ037614 CIT-HSP-2
	39	278.4	9.0	580	102	AQ372558 AQ372558 RPCI11-15
	40	278.4	9.0	547	87	AQ810321 AQ810321 HS_4869_A
	41	274	8.9	906	87	AQ743708 AQ743708 HS_5507_A
	42	274	8.9	608	102	AQ420241 AQ420241 RPCI-11-2
c	43	273.6	8.9	911	87	AQ746594 AQ746594 HS_2278_A
c	44	273	8.9	662	91	B68336 B68336 CIT-HSP-202
c	45	272.2	8.8	658	102	AQ393450 AQ393450 CITBI-E1-

ALIGNMENTS

RESULT 1
AI498192/1
LOCUS AI498192 495 bp mRNA EST 14-APR-1999
DEFINITION tm90g12.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2165446 3', mRNA sequence.
ACCESSION AI498192
VERSION AI498192.1 GI:4390174
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria, Primates, Catarrhini, Hominidae; Homo.
REFERENCE 1 (bases 1 to 495)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948230.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 523 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
FEATURES Location/Qualifiers
source 1..495
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2165446"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pI7I3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTT 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pI7I3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 101 a 122 c 129 g 142 t 1 others
ORIGIN

Query Match 14.2%; Score 427.4; DB 47; Length 495;
Best Local Similarity 98.4%; Pred. No. 5.7e-68;
Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 0;

```
Qy    2176 agtaaatctagaacattctcaaggccc.ttttaatcagatatcaaagccacccccctaaa 2235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     495 AGTAATCTAGAACATTCTCAAGCCCCTTTAATCAGATATCAAAGCCACCCCCCTAAA 436

Qy    2236 gtcaataccctcccaccaaggaatgaatatatgtggacctcactgctcccccaaccccagccct 2295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     435 GTCAATACCTCCCACCAAGGAATGACATATGTGGACCTCACTGCTCCCCCAACCCCAGCCT 376

Qy    2296 caggctgggacacgccaacgctgttccgggttggaacagcacaggctcagaaactgaatg 2355
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     375 CAGGCTGGGACACGCCAACGCTGTTCCGGGTTGGAACAGCACAGGCTCAGAAACTGAATG 316

Qy    2356 tgaaataggcagacctagcaagaggcagatacaggatatcgggcatttgaatatttgaaa 2415
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 TGAAATAGGCAGACTTAGCAAGAGTCAAGATACAGGTATCGGGCATTTGAATATTTGAAA 256

Qy    2416 agtcattcgggaagataaatccagtagctgagcggagc.agctgattcaaagcttaaaa 2475
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 AGTCATTCGGGAAGATAAATCCAGTAGCTGAGCGGAGCAGCTGATTCAAAGCTTAAAA 196

Qy    2476 ccagcgggttcttgttcgggaggcaaatttcctaggaaaaagaagacagacttttctaa 2535
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     195 CCAGCGGGTTCTTGTTCGGGAGGCAAATTTCCTAGGAAAAAGAAGACAGACTTTTCTAA 136

Qy    2536 tgggtccaaatgcggatcactgttagatggactctagaaggactgagctcccttctct 2595
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135 TGTGGTCCAAATGCGGATCACTGTTAGATGGACTCTAGAAGGACTGAGCTCCCTTCTCT 76

Qy    2596 tggaagtatttaagaaaaggctgtgcca 2623
           ||||||||||||||||||||||||||||
Db      75 TGGAAGTATTTAAGAAAAGATCGTGCCA 48
```

RESULT 2
T08680
LOCUS T08680 437 bp mRNA EST 08-AUG-1993
DEFINITION EST06572 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBI25 5' end, mRNA sequence.
ACCESSION T08680
VERSION T08680.1 GI:389708
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Adams,M.D.; Soares,M.B.; Kerlavage,A.R.; Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..437
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):85191"
                     /db_xref="taxon:9606"
                     /clone="HIBBI25"
                     /clone_lib="Infant Brain, Bento Soares"
BASE COUNT      117 a    127 c    107 g     64 t      2 others
ORIGIN
```

```
  Query Match             12.5%;  Score 383.8;  DB 20;  Length 437;
  Best Local Similarity   96.8%;  Pred. No. 1.6e-58;
  Matches  423;  Conservative    0;  Mismatches    9;  Indels    5;  Gaps    3;

Qy    1214 ccatctctcggccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagct 1273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCATCTCTCGCCCTTTCATCAGCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCT 60

Qy    1274 tggaagacctggagcaactcatccctatggaacacgtccagatcccagactgcgtcctgc 1333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGGAAGACCTGGAGCAACTCATCCCTATGGAACACGTCCAGATCCCAGACTGCGTCCTGC 120

Qy    1334 aatacgaagaggaaagactgaaggccaggagggagagcgcgaggccccagccggagtttg 1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATACGAAGAGGAAAGACTGAAGGCCAGGAGGGAGAGCGCGAGGCCCCAGCCGGAGTTTG 180

Qy    1394 tgctgcccaggtctgaagagaagccagaggtggcaccagtggaaaacaggtctgctctgg 1453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCTGCCCAGGTCTGAAGAGAAGCCAGAGGTGGCACCAGTGGAAAACAGGTCTGCTCTGG 240

Qy    1454 tctcagaagatcaggaaacaagcatgtcctgaggcgacgtgagcataacaaaggacatgg 1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTCAGAAGATCAGGAAACAAGCATGTCCTGAGGCGACGTGAGCATAACAAAGGACATGG 300

Qy    1514 aagaagattccagatgcggaaaacctctgtcagacgcccactggccccagatc-tcatc 1572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||| ||||
Db     301 AAGAAGATTCCAGATGCCAGAAAACCTCTNTCAGACGCCCACTGGNCCCAGATCTTCATC 360

Qy    1573 ctgcc-tcatcctgagtcccaatcttccaagggtgccagcccctccgttcatctct---g 1628
           |||||  |||||||||||||||||||||||||||||||||||| |   |    |||   |
Db     361 CTGCCTTCATCCTGAGTCCCAATCTTCCAAGGGTGCCAGCCCCTTCCGTTTCATCTCTGG 420

Qy    1629 aaacccagcatcctttt 1645
           |||||||||||||||||
Db     421 AAACCCAGCATCCTTTT 437


RESULT   3
AI480208/C
LOCUS       AI480208        461 bp    mRNA            EST       14-APR-1999
DEFINITION  tm72f01.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2163675 3'
            similar to contains element LTR1 LTR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI480208
VERSION     AI480208.1  GI:4373376
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   NCI_NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP) Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189284.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna B. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 540  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1..461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2163675"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTTGGAGCGGCGGCGCATAGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia); digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       89 a    112 c    114 g    146 t
ORIGIN
```

```
  Query Match             10.5%;  Score 361.2;  DB 47;  Length 461;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-55;
  Matches  388;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    0;

Qy    2229 caaacccgattaaataggctcgc[illegible]                              2287
           |||||||||||||||||||| [illegible]
Db     461 CAAACACCGTCAATAGCTCGCCA[illegible]                              402

Qy    2288 cccagcgtcaggctgggacacg[illegible]                              2347
           |||||||||||||||||||||||| [illegible]
Db     401 CCCAGCGTCAGGCTGGGACACG[illegible]                              342

Qy    2348 actggctctgaaataggcagac[illegible]                              2407
           |||||||||||||||||||||| [illegible]
Db     341 ACTGGCTCTGAAATAGGCAGACCTAGCAAGAGGAAGATACAGGTATCGGGGCTTTCAGT 282

Qy    2408 ctttggaagtcattcgggaag[illegible]                              2467
           |||||||||||||||||||||| [illegible]
Db     281 GTTTCAAAAGTCATTCGGGAAGATAAATTCATGTCTTGCGGCAGCCAGTTCGATTCAA 222

Qy    2468 agcttggaccagcgggttcttg[illegible]                              2527
           |||||||||||||||||||||| [illegible]
Db     221 AGCTTGGACCAGCGGGTTCTTGTTCGGCAGGCAAATTCCCCTATGAAAAGAAGCAGAAC 162

Qy    2528 ttttctaatgggtccaaatgcg[illegible]                              2587
           |||||||||||||||||||||| [illegible]
```

```
Db    161 IIIICIAATGTGGTCCAAATGCGGATCACIIGCCAGATGGACTCTAGAAGCACTGAGCTC 102

Qy   2599 cctgtctctgtgaaagtatttaagaaaaggctg::cca 2623
          |||||||||||||||||||||||||||| ||| |  ||
Db    101 CCIGTCTCTGGAAGTAIIIAAGAAAAGATCGT3CCA 66


RESULT   4
T05645
LOCUS       T05645          368 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST03534 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDD71, mRNA sequence.
ACCESSION   T05645
VERSION     T05645.1  GI:316795
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nat. Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            9-- Dogger Road, Gaithersbur:, MD 20879
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):82344"
                     /db_xref="taxon:9606"
                     /clone="HFBDD71"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis, lambdaZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT       88 a    100 c     90 g     80 t      4 others
ORIGIN


  Query Match           11.4%;  Score 349.8;  DB 20;  Length 368;
  Best Local Similarity   98.1%;  Pred. No. 1.6e-52;
  Matches  362;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy    984 cgactaccactacatcatggagaacctcttc'gtacgtcatcagcagcttagagctcct 1043
          |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db      1 CGACTACCACTACATCATGGAGAACCTCTTC13TACGTNATCAGCAGCTTAGAGCTCCT 60

Qy   1044 ggtggctaagaactacatgatcgtgtacctgaacggtgccacgcccccggcggaggatgcc 1103
          |||||||||||||||||||||||| ||||||  ||||||||||||||||||||||||||||
Db     61 GGTGGCTGAGGACTACATGATCGTNTACCTGANCGGTGCCACGCCCCCGGCGGAGGATGCC 120

Qy   1104 tggaatcggctggctgaagaagtgctaccagatgatcgaccggaggttgcggaaaaacct 1163
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    121 TGGAATCGGCTGGCTGAAGAAGTGCTACCAGATGATCGACCGGA-GTTGCGGAAAAACCT 179

Qy   1164 gaagtccttgatcatcgtccaccccctcgtggttcattcggactgtggctggccatctctcg 1223
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 GAAGTCCTTGATCATCGTCCACCCCTCGTGG.TCATTCGGACTGTGGCTGGCCATCTCTCG 239

Qy   1224 ccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttggaagacct 1283
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    240 CCCTTTCATCAGCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCTTGGAAGACCT 299

Qy   1284 gaagcaactcatcctatgaaacacgtccaga'cccagactgcgtcctgcaatacgaaga 1343
          |||||||||||||||||||||||||||||'|| |||||||| |||||||||||||||| ||
Db    300 NGAGCAACTCATCCCTATGGAACACGTCCAGA.CCCAGACTTCGTCCTGCAATACGAGGA 359
```

```
Qy   1344 ggaaagact 1352
          |||||||||
Db    360 GGAAAGACT 368


RESULT   5
M78376
LOCUS       M78376          384 bp    mRNA            EST       26-MAY-1992
DEFINITION  EST00524 Fetal brain, Stratagene (cat#936206), Homo sapiens cDNA
            clone HFBCA94, mRNA sequence.
ACCESSION   M78376
VERSION     M78376.1  GI:272692
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
  TITLE     Sequence identification of 2,375 human brain genes
  JOURNAL   Nature 355, 632-634 (1992)
  MEDLINE   92168112
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):77679"
                     /db_xref="GDB:D0S1365E"
                     /db_xref="taxon:9606"
                     /clone="HFBCA94"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT       72 a    127 c     96 g     82 t      7 others
ORIGIN


  Query Match           10.8%;  Score 331.6;  DB 20;  Length 384;
  Best Local Similarity   97.1%;  Pred. No. 2.6e-49;
  Matches  334;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    0;

Qy   1757 cgaggaaggcaagaagcgcaggaggatuaccccgctggcgtcgatggacttcgatgatt 1816
          |||||||||||||||||||||||||| |||||||| |||||||||||||||||||||||
Db      1 CGAGGAAGGCAAGAAGCGCAGGGGGTGGCCCGCNTGGCGTCGGTGGCTTCGGTTTGTT 60

Qy   1817 cgcagcccctgtggtcagagctggatacaagattcaagacccttctcttacttatgaggc 1876
          |||||||||||||||||||| |||||||||||||||||||||||| |||||||||||||
Db     61 CGCAGCCCCTGTGGTCAGAGCTGGATACAAGATTCAAGACCCTTCTNTTGTTTGTNAGCC 120

Qy   1877 gctccaggttggagccacagacacccacggccacccgggctggatctacgtcctttcctg 1936
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTCCAGGTTGGAGCCACAGACACCCACGGCCACCCGGGCTGGATCTACGTCCTTTCCTG 180

Qy   1937 tgcctttccctccagaatgcggcctcagacctagaaagctcaacccccctatgaaaaccag 1996
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCCTTTCCCTCCAGAATGCGGCCTCAGACCTAGAAGCTCAACCCCCCTATGAGGGCAG 240

Qy   1997 gtcctggggtagctcctgacctcctacctatatccaaatttcaacaccatgattttttca 2056
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    241 GTCCTGGGGTAGCTCCTGACCTNCTACCTTATGTCCAAATTTCAACACCATGATTTTTCA 300

Qy   2057 tttgacccgccccctttctgctcataataaaacacccagctcctttt 2100
          ||||||||| |||||||||||||| |||||| | |||| | ||||||
```

```
Db    301 TTTGACCCGGGCCCCIICCCGCICAIAAIGACAACNAGCTICCTT 344


RESULT   6
AQ739838
LOCUS       AQ739838       877 bp    DNA             GSS       16-JUL-1999
DEFINITION  HS_5505_A1_C09_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1081 Col=17 Row=E, genomic survey sequence.
ACCESSION   AQ739838
VERSION     AQ739838.1  GI:5517360
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Feb 19  1999 this sequence version replaced gi:4142171
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1081  row: E  column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 877.
FEATURES             Location/Qualifiers
     source          1. .877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1081 Col=17 Row=E"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      251 a    181 c    242 g    203 t
ORIGIN


  Query Match             10.4%;  Score 321;  DB 82;  Length 877;
  Best Local Similarity   82.1%;  Pred. No. 1.8e-47;
  Matches  395;  Conservative    0;  Mismatches   80;  Indels    6;  Gaps    2;

Qy    2602 tatttaagaaaaggctgggcaggcacgatggctcacgcctgtaatcccag-actttggg 2660
           | | |||   |   |||||||||    | ||| ||| |||| ||||||||| ||||||||
Db      26 TTTCTAAAGCATTACTGGGCCAGATGTGGTGGCTCATGCCTATAATCCCAGCACTTTGGG 85

Qy    2661 aggccgaggcaggcggatcacctgaggtgaggagtttgagaacagcctggccaacatggt 2720
           |||| |||| |||| ||||||||||||| | |||| |||   | ||||| ||||||||
Db      86 AGGCTGAGGTAGGCAGATCACCTGAGGTCAGGAGTTCGAGACAACCCTGGCAACATGGC 145

Qy    2721 gaaacccatctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaa 2780
           ||||||  |||||||||||||||||||||||||| |||| ||| |||||||| ||||| |||
Db     146 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGTGCCTGTAT 205

Qy    2781 tcccagctacttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggttac 2840
           |||||||||| |||||||||||| |||| || ||| |||||| |||||||||||| |
Db     206 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGC 265
```

```
Qy    2841 agtgagccaagatcgtgccactgcactccag...tgggcgacagagcaagactctgtctca 2900
           | |||||||||||| |||  |  |  |  |   |  |  |  |  |  |  |  |  |
Db     266 GGTGAGCCAAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGCAAGACTCTGTCTCA 325

Qy    2901 aaaaaaataaaaaataat-----caggggcacattgctcatgcctgtaatcccagcagcta 2955
           |||||| |||||| |                                          |
Db     326 AAAAAAAAAAAAAAAATTACTGGTTGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTT 385

Qy    2956 tgggaggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggccaacca 3015
           |||||||  ||| | |||| |||| |  |  |  |  |  |  |  |  |  |  |  |
Db     386 TGGGAGGCCAAGGCGAGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACA 445

Qy    3016 tggcgaaaccccgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcct 3075
           ||| |||||||| |||||| | |  |  |  |  |  |  |  |  |  |  |  |  |
Db     446 TGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTACCTGGGCGTGGTGGCACGCACCT 505

Qy    3076 g 3076
           |
Db     506 G 506


RESULT   7
T09124
LOCUS       T09124          335 bp    mRNA            EST       19-AUG-1993
DEFINITION  EST07017 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBP34
            5' end similar to EST containing Alu repeat, mRNA sequence.
ACCESSION   T09124
VERSION     T09124.1  GI:390152
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /db_xref="ATCC [unhosted]:58000"
                     /db_xref="taxon:9606"
                     /clone="HIBBP34"
                     /clone_lib="Infant Brain, Bento Soares"
BASE COUNT      103 a     75 c     89 g     68 t      2 others
ORIGIN


  Query Match             10.1%;  Score 309.8;  DB 20;  Length 335;
  Best Local Similarity   98.2%;  Pred. No. 1.8e-45;
  Matches  323;  Conservative    2;  Mismatches    4;  Indels    2;  Gaps    1;

Qy    2723 aaactcatctctactaaaaatacaaaaattagccagg-gtggtggcaggtgcctgtaa 2781
           | |||||||||||||||||| |  |  |  |  |  | |||||||| ||||| |||
Db       9 AANCTCATCTCTACTAAAAATACA--AAATTAGCCGGGCGTGGTGGCACGCACCTGTAAT 66

Qy    2782 ccagctacttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggttacagt 2842
           ||| |||||||||||||||| |||  |  |  |  |  |  |  |  |  |  |  |  |
Db      67 CCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGT 126

Qy    2843 tgagccaagatcgtgccactgcactccagcctgggcgacagagcaagactctgtctcaaaa 2902
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     127 TGAGCCAAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGCAAGACTCTGTCTCAAA 186
```

```
Qy   2993 aaaaataaaaaataatcaggaacagtgtctatgcctgtaatcccagcactctgggagg 2962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 AAAAATAAAAAATAATCAGGGCACAGTGGCTCATGCCTGTAATCCCAGCACTCTGGGAGG 246

Qy   2963 ctgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggtgaaa 3022
          ||||||||||||||  ||||||||||||||| ||||||||||||||||||||||||||||
Db    247 CTGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGTGAACATGGTGAA 306

Qy   3023 accccgtctctactaaaaatacaaaaatt 3051
          |||||||||||||||||||||||||||||
Db    307 ACCCCGTCTCTACTAAAAATACAAAAATT 335


RESULT   8
AQ588588
LOCUS       AQ588588       590 bp    DNA             GSS       07-JUN-1999
DEFINITION  CITBI-E1-2641J7.TR CITBI-E1 Homo sapiens genomic clone 2641J7,
            genomic survey sequence.
ACCESSION   AQ588588
VERSION     AQ588588.1  GI:5015268
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2641J7"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      190 a    129 c    151 g    120 t
ORIGIN


  Query Match           10.0%;  Score 3 .7.6;  DB 104;  Length 590;
  Best Local Similarity   79.2%;  Pred. No.  4.3e-45;
  Matches  399;  Conservative    0;  Mismatches   84;  Indels   21;  Gaps    2;

Qy   2594 tctggaagtatttaagaaaaggctgggccagtgacgatggctcacgcctgtaatcccag- 2653
          |  .   |  ||   ||  || ||     ||| |||| ||| ||||||||||||||  |
Db     50 TCATGTACTTTAAAAATGAATCATGAGCCCACTACGGTGGCCCACACCTGTAATCCCAGC 109

Qy   2653 actttgggaggccgaggcaggcggatcacctaggtgaggagtttgagaacagcctggcc 2712
          ||||||||||||| |||||||  | ||||  |||||| |||||||  ||| ||||||||||
Db    110 ACTTTGGGAGGCTGAGGCAGATGAATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCC 169

Qy   2713 aacatggtgaaacctcatctctactaaaaatacaaaaattagccaggcgtggtggcaggt 2772
          |||||| |||||| |||||||||||||||||||||||||||||||||||||||||||  |
Db    170 AACATGATGAAACCCCATCTCTACTAAAAATACAAAAATTACCCAGGCGTGGTGGCGCAT 229

Qy   2773 gcctgtaatcccagctacttgggaggctgaggcatgagaatcgcttgaacccgggaggc 2832
          |||||||||||||||||| |||||||||||| || |||| ||| |||| | ||||||
Db    230 GCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGC 289

Qy   2833 gaggttacagtgagccaagatcgtgccactgcactccagcctgggcgacagagcgagact 2892
          ||||||  |||||||| |||||||||||||| |||||||||| ||||||||| | |||
Db    290 GAGGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACT 349

Qy   2893 ctgtctcaaaaaaaaataaaaaa----  -- ------ --t--t---- - --:-:  2932
          |  ||| |||| || | ||||          |  |||  |  |  |||  |||||
Db    350 CCATCTTAAAAATAAAATAAAAAAIAAAIAAAIAAAAIGAATIGGGCIGGGTGTGGTGGC 409

Qy   2933 tcatgcctgtaatcccagcactttgggaggctgaggtgggtggatcacctgaggtcagga 2992
          |||||||||||||||||||| |||| | | ||||| | |||||| |||||||||||||
Db    410 TCATGCCTGTAATCCCAGCACTTTGGGAGACCAAAGCGGGTGGATCACCTGAGGTCAGGA 469

Qy   2993 gttcaagaccagcctggtgaacatggtgaaaccccgtctctactaaaaatacaaaaattag 3052
          |||| ||||||||||    |||| || | ||||||||||||| | |||||| ||||||
Db    470 GTTCGAGACCAGCCTGAGCAACAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTA 529

Qy   3053 gccgggcatggtggtgcatgcctg 3076
          ||| | | ||||||     ||||
Db    530 GCCAGACGTGGTGGCAGGCACCTG 553


RESULT   9
AQ312765
LOCUS       AQ312765       449 bp    DNA             GSS       04-MAY-1999
DEFINITION  RPCI11-1M9.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-1M9,
            genomic survey sequence.
ACCESSION   AQ312765
VERSION     AQ312765.1  GI:4044429
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Adams,M.D., Pounsley,S.D., Zhao,S., Bass,S., Lisher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3555936.
            Other_GSSs: RPCI11-1M9 TVB
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/order.htm) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..449
                     /organism="Homo sapiens"
                     /db_xref="GDB:7500296"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-1M9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      126 a    104 c    131 g     88 t
ORIGIN
```

```
  Query Match             9.9%;  Score 315;  DB 100;  Length 449;
  Best Local Similarity   83.7%;  Pred. No. 1.3e-44;
  Matches  309;  Conservative    0;  Mismatches   60;  Indels    5;  Gaps    1;

Qy    2653 actttgggaggctgaggcaggcggatcac tgaggtgaggagtttgagaaacagcctggcc 2712
           |||||||| || ||||| ||| |||||| | ||||||||||| |||| | ||||| |
Db       4 ACTTTGGGAGGCTGAGGTAGGCAGATCACTTGAGGTGAGGAGTTCGAGACAACCCTGGAC 63

Qy    2713 aacatggtgaaacctcatctctactaaaaatacaaaaattagccaggcgtggtggcaggt 2772
           |||||| | |||||   |||||||||||| | ||||||||||| ||| |||||||| ||
Db      64 AACATGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGT 123

Qy    2773 gcctgtaatcccagctacttgggaggctgaggcatgagaatcacttaaacctgagaggca 2832
           |||| | |||||||||| | |||||| | | ||||||  ||| |||||| ||||||
Db     124 GCCCGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCA 183

Qy    2833 gaggttacagtgagccaagatcgtgccactgcattccagcctgggcgacagagcaagact 2892
           |||||| | ||||||||||||||||||| ||| |||||||||||  |||||| |||||
Db     184 GAGGTTGCGGTGAGCCAAGATCGTGCCACTGCACTCCAGCCTGGGTAACAGAGTGAGACT 243

Qy    2893 ctgtctcaaaaaaaaataaaaaataat-----cagggcacagtggctcatgcctgtaatcc 2947
           |||| |||| |||| |||||| | ||     || | | |||||| || | ||||||
Db     244 CCATCTCAAAAAAAAAAAAAAAAAATTACTGGTTGGGTGCGGTGGCTCACGCTTATAATCC 303

Qy    2948 cagcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcaagaccagcct 3007
           | |||| || | ||  |||  | ||| | ||||||||||||||||||| ||||
Db     304 CAGCACTTTGGGAGGCCAAGGCTAGCGGATCACCTGAGGTCAGGAGTTCAAGACAAGCCT 363

Qy    3008 ggtgaaacatggcgaaaccccgtctctaataaaaatacaaaaattagccgggcatggtggt 3067
           |||| |||| || | |||||  |||||||  | |  ||| ||||||||| |||| ||||||
Db     364 GGCCAACATGGTGAAACCCTGTCTCTACTAAAAATATAAAAATTAGCTGGGCGTGGTGGC 423

Qy    3068 gcatgcctg 3076
           |||| |||
Db     424 AGGCACCTG 432


RESULT  10
AQ215135
LOCUS       AQ215135       511 bp    DNA             GSS       19-SEP-1998
DEFINITION  HS_3249_B1_E06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate 3249 Col=11 Row=J, genomic survey
            sequence.
ACCESSION   AQ215135
VERSION     AQ215135.1  GI:3626336
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3249  Row: J  Column: 11
            Class: BAC ends
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1..511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate 3249 Col=11 Row=J"
                     /clone_lib="CIT Appr. Hum. Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      147 a    122 c    102 g    139 t      1 others
ORIGIN


  Query Match             9.8%;  Score 311.8;  DB 99;  Length 511;
  Best Local Similarity   80.5%;  Pred. No. 4.3e-44;
  Matches  376;  Conservative    1;  Mismatches   85;  Indels    6;  Gaps    2;

Qy    2616 ctgggccaggcacgatggctcacggcctgtaat ccccagcactttgggaggccgaggcg 2674
           |||| |||||||||||||||||||||||||||| |||| |||||||||||||||||||
Db      43 CTGNGCCAGATGTCGTCGCTCATCCCTAATAAT CCCNGCACTTTGGGAGGCCGAGGCAG 102

Qy    2675 ggatcacctgaggtgaggagtttgagaccagc tggccaacatggtgaaaccctgtctct 2734
           |||| |||||||||| |||||| | |||||||  |||| |||||||||| |||| ||||
Db     103 AGATCACCTGAGGTGAGGAGTTCGAGACAACT CGGCAACATGGTGAAACCCTGTCTCT 162

Qy    2735 actaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctacttgg 2794
           |||| |||||||| |||| |||| | |||| | |||| |||| |||| |||| ||||||
Db     163 ACTAAAAATACAAAAATTAGCCGGGCATGGTGGCGCGTGCCTGTAATCCCAGCTACTCGG 222

Qy    2795 gaggctgaggcatgagaatcacttaaacctgaggcggaggttgcagtgagccgagatcgt 2854
           |||| |||||||| |||| | |||||||| | |||| |||| |||| |||| | ||||||
Db     223 AAGGCTGAGGCAGGAGAATCGCTTGAACCTG GGAGGCAGAGTTGCAGTGAGCCAAGATCGC 282

Qy    2855 gtgccactgcattccagcctgggcgacagag caagactctgtctcaaaaaaaaaaaaaaa 2914
           |||| |||| |||| |||| |||| |||| | |||| |||| |||| |||| ||||||||
Db     283 GTGCCACTGCACTCCAGCCTGGGCAACAGA GCGAGACTCTGTCTCAAAAAAAAAAAAAAA 342

Qy    2915 taat-----cagggcacagtggctcatgcctgtaatcccagcactttgggaggctgaggct 2969
           ||||     ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db     343 AATTACTGGTTGGGTGCGGTGGCTCACGCCTATAATCCCAGCACTTTGGGAGGCCGAGGCG 402

Qy    2970 gggtggatcacctgaggtcaggagttcaagaccagcctggtgaaacccgtctctactaaaa 3029
           |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||||
Db     403 TAGCGGATCGCCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTC 462

Qy    3030 ctctaataaaaatacaaaaattagccgggc atggtggtgcatgcctgtaatcccagctac 3089
           |||| |||||||| |||| |||| |||| | |||| |||| |||| |||| |||| ||||
Db     463 CTCTAACTAAAAAATATAAAAATTAGCTGG GCGTGGTGGCACGCCTGTAATCCCAGCTAC 522
```

```
RESULT  11
AQ381918/c
LOCUS       AQ381918       660 bp    DNA             GSS       21-MAY-1999
DEFINITION  RPCI11-135N16.TV RPCI-11 Homo sapiens genomic clone RPCI-11-135N16,
            genomic survey sequence.
ACCESSION   AQ381918
VERSION     AQ381918.1  GI:4352941
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Deslattes,E. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:2...
            Other_GSSs: RPCI11-135N16.TC
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 2...
            Tel: 301 838 0200
            Fax: 301 838 0208
```

```
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..660
                     /organism="Homo sapiens"
                     /db_xref="GDB:7551783"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-135N16"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      136 a    180 c    152 g    192 t
ORIGIN


  Query Match             9.7%;  Score 298.2;  DB 102;  Length 660;
  Best Local Similarity   80.1%;  Pred. No. 1.3e-43;
  Matches  378;  Conservative    0;  Mismatches   88;  Indels    6;  Gaps    2;

Qy    2610 aaaaggctgggcaggcacggatggctcacgcctgtaatcccag-actttgggaggccgag 2668
           |||||  ||||| |||||| || ||  |||||||||| ||| | | ||||| ||
Db     660 AAAATAATGGGCTGGGCACGGTGGCTCATGCCTATAATCCCAGCACTGTTGAAGGCCAAG 601

Qy    2669 gcaggcggatcacctgaggtcaggagttcgagaccagcctggccaacatggtgaaaccctg 2728
           |||||||||| |||||| |||||||  | || ||||||||||||||||| |||||| |
Db     600 CGGGGCGGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCGAAACCCC 541

Qy    2729 atctctactaaaaatacaaaaattagccaggcatggtggcaggtgcctgtaatcccagct 2788
           ||| ||| |||'||| |||| |||||||  ||  || ||||||| |||| ||||||||||
Db     540 ATCCCTAGTAAAAATGCAAAGATTAGCTGGGTATGTTGGCAGGCGCCTATAATCCCAGCT 481

Qy    2789 actttgggaggctgaggcatgagaatcacttgaacctgaggaggcagaggttacagtgagcc 2848
           |||  | ||| |||  | |||||| | |   |||| |||:|||| ||| ||||||||||
Db     480 ACTCAGGAGGCTGAGATAGGAGAATCGCTTGAACCTGGGAGGCAGAAGTTGCAGTGAGCC 421

Qy    2849 aagatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaat 2908
            | || | | ||||||  || ||| |||' ||| ||||| ||||   ||||||| || |
Db     420 CAAATAGGGTCACTGCGCTCTAGCTTGGGAGACGGAGCACGACTTCATCTCAAAGAATAG 361

Qy    2909 aaaaaataat.....caggcacagtggctcatgcctgtaatcccagcactctgggaggc 2963
           ||| | |      | |||| ||||| ||  |||||||||||||||| |||| | |
Db     360 TAATAATAATGTTGGCCGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTGTGGGAAGT 301

Qy    2964 tgaggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaa 3023
           |||||||| |||||||| |||||||  |||||||||||  |||| || ||||
Db     300 CAAGGTGGGTGAGTCACCTGAGCACAGGAGTTGAGACCAGCCTGGCCAACAGGGTGAAA 241

Qy    3024 ccccatctctaataaaaatacaaaaattagccgggcatggtggtgcatgcct 3075
           |||| |||||| ||||.|||||||||| ||  |  |||||||||| |||||
Db     240 CCCCATCTCTACTAAAAATACAAAAATTAGCCGGGTATATGGTGGTGCGTGCCT 189


RESULT  12
AQ748222
LOCUS       AQ748222       922 bp    DNA             GSS       19-JUL-1999
DEFINITION  HS_5536_A1_G08_SP6 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1112 Col=15 Row=M, genomic survey sequence
ACCESSION   AQ748222
VERSION     AQ748222.1  GI:5295380
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 822)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holmes,I.,
            Keller,A., Shaker,P., Furlong,T., Young,T., Plaetke,R., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.htm)
            or from Research Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1112  row: M  column: 15
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 822.
FEATURES             Location/Qualifiers
     source          1..822
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1112 Col=15 Row=M"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      241 a    173 c    207 g    199 t      2 others
ORIGIN


  Query Match             9.7%;  Score 298;  DB 87;  Length 822;
  Best Local Similarity   80.0%;  Pred. No. 2.1e-43;
  Matches  396;  Conservative    0;  Mismatches   76;  Indels   23;  Gaps    3;

Qy    2605 ttaagaaaaaggctgggcaggcacggatggctcacgcctgtaatcccag-actttgggagg 2663
           || |  ||    |||||| ||| |  || |||| || ||||||| |||.|  |||||
Db      92 TTTAAGAATTAATGGGCCGGGCGCAGTGACTCATGCCTTGTAATCCCAGCACTTTGGGAGG 151

Qy    2664 ccgaggcaggcggatcacctgaggtgaggagtttgagaccagcctggccaacatggtgaaa 2723
           ||||||| || |||||||||||| ||||||||||| |||||  |||||  |||| ||||||
Db     152 CCGAGGCGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCCAGCCAACGTGGTAAA 211

Qy    2724 acctcatctctactaaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatccc 2783
           ||| ||||||||||||||||||| |||||| ||| |||| |  |||| |||||||||||
Db     212 ACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGTGCATGCCTGTAATCC 271

Qy    2784 cagctacttgggaggctgaggcatgaggaatca-cttaaacctgaggaggcagaggttacag 2842
           |||||||| |||| ||||| ||| |||||  | ||| |||||| ||||| | | | |||
Db     272 CAGCTACTGGGGATGCTGAAGCAGGAGAAGAATCTTGAACCTGGGAGGCGGAGCTTGCAG 331

Qy    2843 tgagccaagatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaa 2902
           |||||| |||| | |||||||| ||||||||||||||||| |||||  |||| |||||||
Db     332 TGAGCCGAGATGGCGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCTGTCTCAAA 391

Qy    2903 aaaaataaaaaataa----------------------tcagggcacagtggctcatgcctg 2941
           ||||| ||||||  ||                      |.|||  |||| | | |||
Db     392 AAAAAAAAAAAAAAAAAAAAAGAATTAAATGATGGAGGCTTGTGTGGTGGCGCACGCCTG 451

Qy    2942 taatcccagcactctgggaggctgaggtgggtggatcacctgaggtcaggagttcgagacc 3001
           ||| |||||||| ||||||||  |||||| ||||||| | |
Db     452 TAACCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGAC 511

Qy    3002 cagcctggtgaacatggcgaaaccccgtctctaataaaaatacaaaaattag.ccgggcat 3061
           |||||||  |||||| |||||||| |||||||||||||||||||||||||| | ||||
Db     512 CAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCGT 571
```

```
Qy   3062 ggtggtgcatgcctg 3076
          ||||||||||||||| 
Db    572 GGTGGTGCATGCCTG 586


RESULT  13
AQ346434
LOCUS       AQ346434      705 bp    DNA             GSS       07-MAY-1999
DEFINITION  RPCI11-101K7 TJB RPCI-11 Homo sapiens genomic clone RPCI-11-101K7,
            genomic survey sequence.
ACCESSION   AQ346434
VERSION     AQ346434.1  GI:4171330
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-101K7.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Homo sapiens"
                     /db_xref="GDB:7536646"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-101K7"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      212 a    154 c    195 g    144 t
ORIGIN


  Query Match             9.6%;  Score 206.6;  DB 102;  Length 705;
  Best Local Similarity   80.8%;  Pred. No. 3.7e-43;
  Matches  374;  Conservative    0;  Mismatches   79;  Indels   10;  Gaps    2;

Qy   2607 aagaaaaggctgggccaggcacgatggctcacgcctgtaatcccag-actttgggaggcc 2665
          ||   ||  |  ||||||| |  ||||| |||||||||||||||| ||||||||||||
Db     20 AATTCAAATCCAGGCCAGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCT 79

Qy   2666 gaggcaggcggatcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaac 2725
          |||||||| |||||||||||| ||||| | ||| ||||||||||||||||| ||||
Db     80 GAGGCAGGAGAATCACCTGAGGTTAGGAGGTCAAGACCAGCCTGGCCAACATGGCAAAAC 139

Qy   2726 ctcatctctactaaaaatacaaaaattagctgggcgtggtggcaggtgcctgtaatccca 2785
          | | | ||||||||||||||||||||||||||||||||   || ||||||||  ||||||||||||
Db    140 CCCGTGTCTACTAAAAATACAAAAAATTAGCCGGGCATGGTGGCACACGCCTGTAATCCCA 199

Qy   2786 gctacttgggaggctgaggcatgagaatcacttaaacctgagaggcagaggttacagtga 2845
          |||||||||||||||||||| ||||||| | ||||||  |||||||||||| ||||||
Db    200 GTTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGA 259

Qy   2846 gccaagatcgtgccactgcattccagcctgggcgacagagcgagactctctctcaaaaaaa 2905
          ||| |||| |  |||||| | |  |  | |  |  |  |  |  |  |  |  |  |
Db    260 GCCGAGATGGCACCACTGCACTTCAGCCTGGGCAACAGAGCGAGACTCCATCTCAAAAAA 319

Qy   2906 aataaaaaa---------taatcagggcacagtggctcatgcctgtaatcccaagcactt 2956
          || ||||||         ||  |  |  |  |  |  |  |  |  |  |  |  |
Db    320 AAAAAAAAATTCAGATCCTAAGTTGGGTGCAGTGGCTCATGCCTATAATCCCAGCACTT 379

Qy   2957 gggaggctgaggtgggtggatcacctgaggtcaggagttcgagaccagcctggccaacat 3016
          |||||||||||||  ||  ||||  |  |  |  |  |  |  |  |  |  |  |
Db    380 GGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGAT 439

Qy   3017 ggcgaaaccccgtctctaaaaaaatacaaaaattagccgggcg 3059
          |  ||  ||| | |||  || | |  |  |  |  |  |  |
Db    440 AGAGAGACCTCATCTCAAAAAAAAAAAAAAAAAAAAAAAGGTC 482


RESULT  14
B68229/c
LOCUS       B68229        558 bp    DNA             Jun       18-JUN-1998
DEFINITION  CIT978SK-A-616D9.TV.1 CIT978SK Homo sapiens genomic clone A-616D9,
            genomic survey sequence.
ACCESSION   B68229
VERSION     B68229.1  GI:2666939
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Kerlavage,A.R.,
            Golden,K., Berry,K., Granger,L., Sun,H., Wible,C., Kim,U.-J.,
            Shizuya,H., Simon,M. and Venter,J.C.
  TITLE     Use of a human BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CIT978SK-A-616D9.1; CIT978SK-A-616D9.TV
            CIT978SK-A-616D9.TP.1
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /db_xref="GDB:1342711"
                     /db_xref="taxon:9606"
                     /clone="A-616D9"
                     /clone_lib="CIT978SK"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
                     CalTech Human BAC Library A"
BASE COUNT       96 a    163 c    143 g    156 t
ORIGIN


  Query Match             9.6%;  Score 206;  DB 81;  Length 558;
  Best Local Similarity   82.8%;  Pred. No. 4.8e-43;
  Matches  389;  Conservative    0;  Mismatches   70;  Indels   11;  Gaps    4;

Qy   2616 ctgggcgaggcacgatggctcacgcctgtaatcccagcactttgggaggccgagg 2674
          |||||||||  || |||||||  |  |  |  |  |  |  |  |  |  |  |
Db    542 CAGGGCCAGGAGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGG 483
```

```
Qy    2675 ggatcacctgaggtgaggagtttgagaacagcctggccaacatggtgaaacctcatctct 2734
           ||||| |||||||||| |||||||||| ||| ||||||| || ||||||| ||||||||||
Db     482 ACATCAC--GAGGTCAGGAGACTGAGACCAGCCTGGCTAACATGGTGAAACCCCATCTCT 425

Qy    2735 actaaaaatac----aaaaattagccaggcgtggtggcaggtgcctgtaatcccagctac 2790
           ||||||||||    |||||||||| ||| ||||||| || ||||||| ||||||||||
Db     424 ACTAAAAATACAAAAAAAAAATTAGCCGGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTAC 365

Qy    2791 ttgggaggctgaggcatgagaatcacttaaacctgaggaggcagaggttacagtgagccaa 2850
           | ||||||||||||| |||||| | | ||  | |||| ||| || ||||||||| |
Db     364 TCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGTGGAGCTTGCAGTGAGCCGA 305

Qy    2851 gatcgtgccactgcattccagcctgggcgacagagcaagactctgtctcaaaaaaaataa 2910
           |||| |||| || | |||||||| ||  ||| |||||  ||||||||||||| ||
Db     304 GATCGGGCCACGGCACTCCAGCCTGGGTGACAGAGCGAGACTCCATCTCAAAAAAAAAA 245

Qy    2911 aaaataatca----gggcacagtggctcatgcctgtaatcccagcactctgggaggctga 2966
           ||| | |||     ||||| |||| |||||||||||||| ||||||| |||
Db     244 AAAGTCATCAGGCTGGGCATGGTGGCACATGCCTGTAATCCCAGCACTTTGGGAGGGTGA 185

Qy    2967 ggtgggtggatcacctgaggtcaggagttcaagaccagcctggtgaacatggcgaaaccc 3026
           ||   | |||||||||||||||||||| ||||||||||| |||||||| ||||||
Db     184 GGCCAGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGTCAACATGGCAAAACCC 125

Qy    3027 cgtctctaataaaaatacaaaaattagccgggcatggtggtgcatgcctg 3076
           || | | ||||||||||||||||| | |  |||||   ||||||
Db     124 CGACTCTACTAAAAATACAAAAATTATCCAGGTGTGGTGGCAGGTGCCTG 75


RESULT  15
AA601278/c
LOCUS       AA601278        440 bp    mRNA            EST       01-OCT-1997
DEFINITION  no15b09 s1 NCI_CGAP_Phe1 Homo sapiens cDNA clone IMAGE:1100729 3'
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA601278
VERSION     AA601278.1  GI:2434903
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637872.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 657  Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1100729"
                     /clone_lib="NCI_CGAP_Phe1"
                     /tissue_type="pheochromocytoma"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Pheochromocytoma. 5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'
                     Average insert size:  1.3 kb."
BASE COUNT       86 a    123 c    100 g    131 t
ORIGIN


  Query Match             9.5%;  Score 292;  DB 36;  Length 440;
  Best Local Similarity   84.0%;  Prod. No. 2.5e-42;
  Matches  353;  Conservative    0;  Mismatches   65;  Indels    2;  Gaps    2;

Qy    2619 ggccaggcacgatggctcacgcctgtaatcccag-actttgggaggccgaggcgggcgga 2677
           ||| |   | |||||||| |||||||||||||| ||||||||| | |||| ||| |||
Db     420 GGCTGGATGCAATGGCTCATGCCTGTAATCCCAGCACTTTGGGAGACTGAGGCAGGTGGA 361

Qy    2678 tcacctgaggtgaggagtttgagaccagcctggccaacatggtgaaacctctctactaaa 2737
           |||| |||| | || |||||| || ||||||||||||||||||||||| |||| |||
Db     360 TCACTTGAGATCAGCAGTTTGTGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACT 301

Qy    2738 aaaaatacaaaaattagccaggcgtggtggcaggtgcctgtaatcccagctactcgggag 2797
           |||||||||||||||||||||||| |||||||| |||||||||||||| | |||| |||
Db     300 AAAAATACAAAAATTAGCCAGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTCAAGAG 241

Qy    2798 gctgaggcatgagaatcacttaaacctgaggaggcagaggttacagtgagccaagatcgtg 2857
           |||||||||  |||||| ||| ||   | |||| |||||| ||||||||  |||||
Db     240 GCTGAGGCAGAAGAATCGCTTGAAACTGGGAGGTGGAGGTTGCAGTGAGCTGAGATCGCA 181

Qy    2858 ccactgcattccagcctgggcgacagagcaagactctgtctca-aaaaaaataaaaaata 2916
           ||| |||| | |||| |||||||||||| |||||| |||||| |||||| ||||||
Db     180 CCATTGCACTACAGCTTGGGCGACAGAGCGAGACTCCGTCTCAGAAAAAAAAAAAAAAAA 121

Qy    2917 atcagggcacagtggctcatgcctgtaatcccagcactctgggaggctgaggcgggcgga 2976
           | |||||||||| |||  |||||||||||||||| |||||||||||| || |
Db     120 GGCTGGGCACAGTGGTTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGA 61

Qy    2977 tcacctgaggtcaggagttcaagaccagcctggccaacatggtgaaaccccgtctctaaat 3036
           |||| |||| |||||||||||||||||||||  |||||| | |||||| ||||| |
Db      60 TCACTTGAGTTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACT 1


Search completed: February  7, 2000, 19:32:30
Job time: 6660 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 02:58:19 ; Search time 159.56 Seconds
(without alignments)
701.293 Million cell updates/sec

Title: US-09-174-937-6
Perfect score: 966
Sequence: 1 atggaggagggagacagagtt..... tgaagctgaaagaaaagcct 966

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA.*

Word size : 6

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	284.8	29.5	1386	3	US-08-408-095-22	Sequence 22, Appl
c 2	42.6	4.4	7218	1	US-08-282-463-14	Sequence 14, Appl
c 3	32	3.4	777	4	US-09-975-316-37	Sequence 37, Appl
c 4	32.8	3.4	5495	2	US-08-662-010A-1	Sequence 1, Appli
5	32.8	3.4	5495	2	US-08-662-010A-2	Sequence 2, Appli
c 6	32.8	3.4	1248	2	US-08-662-010A-9	Sequence 9, Appli
c 7	32.8	3.4	5495	2	US-08-680-726A-1	Sequence 1, Appli
8	32.8	3.4	5495	2	US-08-680-726A-2	Sequence 2, Appli
c 9	32.8	3.4	1248	2	US-08-680-726A-9	Sequence 9, Appli
c 10	32.8	3.4	10592	2	US-08-680-726A-51	Sequence 51, Appl
11	32.8	3.4	10592	2	US-08-680-726A-52	Sequence 52, Appl
12	31.8	3.3	2277	2	US-08-675-967-2	Sequence 2, Appli
13	31.8	3.3	2277	2	US-08-675-974-2	Sequence 2, Appli
14	31.8	3.3	2277	4	US-09-009-487-2	Sequence 2, Appli
15	31.6	3.3	1357	6	5340934-7	Patent No. 5340934
16	31.4	3.3	2659	2	US-08-771-609A-1	Sequence 1, Appli
17	30.6	3.2	276	1	US-07-901-270-1	Sequence 1, Appli
18	30.6	3.2	276	1	US-08-057-168-1	Sequence 1, Appli
19	30.4	3.1	4080	1	US-08-375-300-1	Sequence 1, Appli
20	30.4	3.1	2295	1	US-08-375-300-3	Sequence 3, Appli
21	30.4	3.1	994	2	US-08-817-997A-1	Sequence 1, Appli
22	30.4	3.1	4080	5	PCT-US95-16930-1	Sequence 1, Appli
23	30.4	3.1	2295	5	PCT-US95-16930-3	Sequence 3, Appli
c 24	30	3.1	1737	1	US-08-202-056-4	Sequence 4, Appli
c 25	30	3.1	1737	1	US-08-375-093A-3	Sequence 3, Appli
26	30	3.1	7573	1	US-08-297-959-2	Sequence 2, Appli
27	30	3.1	3204	1	US-08-420-935-1	Sequence 1, Appli
c 28	30	3.1	1317	2	US-08-153-848-45	Sequence 45, Appl
c 29	30	3.1	1737	2	US-08-701-265-3	Sequence 3, Appli
c 30	30	3.1	1737	3	US-08-268-586-3	Sequence 3, Appli
c 31	30	3.1	1737	3	US-08-805-478-3	Sequence 3, Appli
c 32	30	3.1	1737	3	US-08-904-627A-3	Sequence 3, Appli
c 33	30	3.1	1737	4	US-08-901-238-3	Sequence 3, Appli
c 34	30	3.1	1737	4	US-08-901-227-3	Sequence 3, Appli
c 35	30	3.1	1317	5	PCT-US93-11178-45	Sequence 45, Appl
c 36	30	3.1	1737	5	PCT-US94-09894-69-2	Sequence 2, Appli
37	30	3.1	4393	6	5175295-4	Patent No. 5175295
38	30	3.1	4393	6	5177307-1	Patent No. 5177307
39	29.8	3.1	1759	2	US-08-472-520-5	Sequence 5, Appli
40	29.8	3.1	1759	2	US-08-487-994-5	Sequence 5, Appli
41	29.8	3.1	1759	4	US-08-472-520A-5	Sequence 5, Appli
42	29	3.0	6605	2	US-08-705-381A-4	Sequence 4, Appli
c 43	29	3.0	2400	2	US-08-705-085-1	Sequence 1, Appli
c 44	29	3.0	2400	4	US-08-921-581-1	Sequence 1, Appli
c 45	28.9	3.0	3395	6	5221421-1	Patent No. 5229424

ALIGNMENTS

RESULT 1
US-08-408-095-22
; Sequence 22, Application US/08/408,095
; Patent No. 5858678
; GENERAL INFORMATION:
; APPLICANT: Chinnadurai, Govindaswamy
; TITLE OF INVENTION: APOPTOSIS-REGULATING PROTEINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak, Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,095
; FILING DATE: 21-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mack, Susan J.
; REGISTRATION NUMBER: 30,951
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 214..1185
US-08-408-095-22

Query Match 29.5%; Score 284.8; DB 3; Length 1386;
Best Local Similarity 62.9%; Pred. No. 2.3e-80;
Matches 496; Conservative 0; Mismatches 277; Indels 0; Gaps 0;

Qy 177 gagaattcaagctcacagcccccctctat aat ttt t ttt. 236

Db 368 CAGAAACAAACTAATGGGTGCCAACATAAGAGCA.A TG.A. TTA.T.A. 419

Qy 237 tctctctgatgataaattggaatttcttttttt.aaattgtttt ttttt 296

Db 420 ATTGTCAGATGATTTGGATGAAA---GGTTGATTTAGTTAATTTTTTA---TAT 473

Qy 297 tagcgggatgaagcagattctt tgatagct tgtt gtttt aa ttttt 356

```
Db    474 ACCGTCAGAGAATAGTAATGAGTTTGAGTGGAAGATGATCTTCCAAAACCCAAGACTAC 533

Qy    357 ttctggccaagagtcagagtctattccagaatatacggccgaagaggaacgggaggacaa 416

Db    534 TGAAGTAATTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGAAGATGG 593

Qy    417 cggctttggatgacagtggtcattggaaaacaagagcagccgcattgacatgaaggtcat 476

Db    594 ACGACGTTGTGTTATGTTCAGGATTGGAGAATAGGACCACAGGTTGATATGAAGGTAAT 653

Qy    477 cgaggccctacaggagagtcatttctccacggaaggagattcaggatactatgggggacggtct 536

Db    654 TGAACCCTATAAAAAAGTTATCAGCCATGG---------GGGATATTATGGGGATGGATT 704

Qy    537 aaatgccatcattgtgtttgccgcctgttttgtgccagacagcagtcgggcggattacca 596

Db    705 AAATGCCATTGTTGTATTTGCTGTCTGTTTCA GCCTGAAAGTAGTCAGCCTAACTATAG 764

Qy    597 ctatgtcatggaaaatcttttcctatatgtaataagtactttagagttgatggtagctga 656

Db    765 ATACCTGATGGACAATCTTTTTAAATATGTTATTGGCACTTTGGAGCTATTAGTAGCAGA 824

Qy    657 agactatatgattgtgtacttgaatggtgcaacccaagaaggaggatgccagggctagg 716

Db    825 AAACTACATGATAGTTTATTTAAATGGTGCAATAACTGGAACAAAAATGCCCAGTCTGGG 884

Qy    717 ctggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttgaaatcatt 776

Db    885 ATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAAAAATCTAAAATCCCT 944

Qy    777 catcattgttcatccatcttggttcatcagaaaaatccttgttgtgacagcagccttttat 836

Db    945 AATCATTGTACATCCTTCTTGGTTTATCAGAATACTTCTGGCTGTTACAAGACCATTTAT 1004

Qy    837 aagttcaaaattcagcagtaaaattaaatatttcaatagcttatcagaaactcagtgggct 896

Db   1005 TAGCTCGAAATTCAGCCAAAAAATTAGATACGTGTTTAATTTGGCAGAACTAGCAGAACT 1064

Qy    897 gatgctgatggattgatcatgattccaatggtcatcatcaatattgacttgaaagctgaa 956

Db   1065 TGTCCCATGGAATACGTTGGCATACCAGAATCATAAAACAAGTTGATCAAGAACTTAA 1124

Qy    957 agaaaag 964

Db   1125 TGGAAAAG 1132


RESULT   2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT POXVIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33472/214 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899119
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS.
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.4%;  Score 42.6;  DB 1;  Length 7218;
  Best Local Similarity   5.1%;  Pred. No. 0.0015;
  Matches   21;  Conservative  214;  Mismatches  178;  Indels    0;  Gaps    0;

Qy    358 tctggccaagagtcagagtctattccagaatatacggccgaagaggaacgggaggacaa 417

Db   1472 TGTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACGPPPPPPPPPPPPPPPPPPPP 1413

Qy    418 cggctttggatgacagtggtcattggaaaacaagagcagccgcattgacatgaaggtcat 477

Db   1412 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1353

Qy    478 gaggccctacaggagagtcatttctccacggaaggagattcaggatactatgggggacggtct 537

Db   1352 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1293

Qy    538 aatgccatcattgtgtttgccgcctgttttgtgccagacagcagtcgggcggattacca 597

Db   1292 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1233

Qy    598 tatgtcatggaaaatcttttcctatatgtaataagtactttagagttgatggtagctga 657

Db   1232 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1173

Qy    658 gactatatgattgtgtacttgaatggtgcaacccaagaaggaggatgccagggctagg 717

Db   1172 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1113

Qy    718 tggatgaagaaatgctaccagatgattgacagacggttgaggaagaatttgaaatcatt 777

Db   1112 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1053


RESULT   3
US-08-975-316-37/c
; Sequence 37, Application US/08975316
; Patent No. 5952486
;  GENERAL INFORMATION:
;    APPLICANT:  BLOKSBERG, Leonard N.; HAYUKKALA, Tikka
;    APPLICANT:  and GRIERSON, Alastair W.
;    TITLE OF INVENTION:  MATERIALS AND METHODS FOR
;    TITLE OF INVENTION:  THE MODIFICATION OF PLANT LIGNIN CONTENT
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
```

```
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/975,316
;       FILING DATE:
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/713,000
;       FILING DATE:  September 11, 1996
;     ATTORNEY/AGENT INFORMATION
;       NAME:  SLEATH, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000/[illegible]-C1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-269-0565
;       TELEFAX:  206-269-0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  37:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  777 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-975-316-37


  Query Match             3.4%;  Score 33;  DB 4;  Length 777;
  Best Local Similarity   49.7%;  Pred. No. 0.47;
  Matches   84;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     260 gtccagatgaaaattgacatcaatgtggatgaa[illegible]ttgatac[illegible]gatgaaggcagattttt 319
           |||| ||| |       ||||| |   |||  | ||       |  |||||||  |||
Db     469 GTCCTCATGATCAAAATGATCAACATTTTGGAAGATCACTTGGAGGTGGAAGCAGGGTCTC 410

Qy     320 ttgagtacactggccatgatcccacaggccaacaaagattctggccaagagtcagagtcta 379
           || ||  | ||  |||| ||| ||        |  ||| | ||   |||      |
Db     409 TTTAGGCCCTTTGCCCATTGTCCACCAAGAAATGCCATCTTCAGTCCCCAAGTTGTGTGCAC 350

Qy     380 ttccagaatatacggccgaagaggaacgggaggacaaccggctttggat 428
           |||||||    |  || | |   |||   |||    | ||| || ||
Db     349 TTCCAGATGGCAGTGCAGGAACCAAACTCCTGGATTGTCTGCTGTGAAT 301


RESULT   4
US-08-602-010A-1/c
; Sequence 1, Application US/08602010A
; Patent No. 5753235
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,010A
;      FILING DATE:  February 15, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,720
;      REFERENCE/DOCKET NUMBER:  2618-49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303/863-9700
;      TELEFAX:  303/863-0223
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5495 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-602-010A-1


  Query Match             3.4%;  Score 32.8;  DB 2;  Length 5495;
  Best Local Similarity   53.0%;  Pred. No. 1.7;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaaatttgaaatcattcatc[illegible]ttttcatccgtcttt 797
           || |  ||| ||||  ' '|  |  |  |  |  |  |  |  |
Db    4619 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATCATCAC 4560

Qy     798 gttcatcagaacaatcctttgctgtgacaaga[illegible]ttttataagttcaaaattcaaaatgctgtga 857
           ||| || |||     ||  |                                    |
Db    4559 GTTGTTCTGAAATTGATTTACCCAAACTAGCAACAAAAAGTTGAGAAGTTGTAGTCA 4500

Qy     858 aattaaatatgt 869
           |||'|||  | |
Db    4499 AATTAAAAGTCT 4488


RESULT   5
US-08-602-010A-2
; Sequence 2, Application US/08602010A
; Patent No. 5753235
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,010A
;      FILING DATE:  February 15, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,720
;      REFERENCE/DOCKET NUMBER:  2618-49
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303/863-9700
;      TELEFAX:  303/863-0223
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5495 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

US-08-602-010A-2

Query Match 3.4%; Score 32.8; DB 2; Length 5495;
Best Local Similarity 53.0%; Pred. No. 1.7;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
Db 877 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 936

Qy 798 gttcatcagaacaatccttgctgtgacacgaccttttataagttcaaaattcagcagtaa 857
Db 937 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 996

Qy 858 aattaaaatatgt 869
Db 997 AATTAAAAGTCT 1008

RESULT 6
US-08-602-010A-9/c
; Sequence 9, Application US/08602010A
; Patent No. 5753235
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,010A
; FILING DATE: February 15, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1248
US-08-602-010A-9

Query Match 3.4%; Score 32.8; DB 2; Length 1248;
Best Local Similarity 53.0%; Pred. No. 0.72;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
Db 922 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 863

Qy 798 gttcatcagaacaatccttgctgtgacacgaccttttataagttcaaaattcagcagtaa 857
Db 862 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 803

Qy 858 aattaaaatatgt 869
Db 802 AATTAAAAGTCT 791

RESULT 7
US-08-680-726A-1/c
; Sequence 1, Application US/08680726A
; Patent No. 5804197
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,726A
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-CI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5495 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-680-726A-1

Query Match 3.4%; Score 32.8; DB 2; Length 5495;
Best Local Similarity 53.0%; Pred. No. 1.7;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
Db 4619 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 4560

Qy 798 gttcatcagaacaatccttgctgtgacacgaccttttataagttcaaaattcagcagtaa 857
Db 4559 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 4500

Qy 858 aattaaaatatgt 869
Db 4499 AATTAAAAGTCT 4488

RESULT 8
US-08-680-726A-2
; Sequence 2, Application US/08680726A
; Patent No. 5804197
; GENERAL INFORMATION:

```
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-11
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5495 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-726A-2


  Query Match             3.4%;  Score 32.8;  DB 2;  Length 5495;
  Best Local Similarity   53.0%;  Pred. No.: 1.7;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
           |:|:|| |:|:|:|:|:|:|:|:||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     877 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 936

Qy     798 gttcatcagaacaatccttgctgtgcgacagca.ttttataa;ttcaaaatttcagcagtaa 857
           ||| ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     937 GTTGTTCTGAAATTGATTTACCCAAACTACCAATCAATAAACTTTGAGAACCTGCACCCA 996

Qy     858 aattaaaatatgt 869
           |:|:|:|:|:|:|
Db     997 AATTAAAAGTCT 1008


RESULT  9
US-08-680-726A-9/c
; Sequence 9, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-11
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1248 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1248
US-08-680-726A-9


  Query Match             3.4%;  Score 32.8;  DB 2;  Length 1248;
  Best Local Similarity   53.0%;  Pred. No.: 1.7;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaagaatttgaaatcattcatcattgttcatccatcttg 797
           |:|:|| |:|:|:|:|:|:|:|:||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     922 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 863

Qy     798 gttcatcagaacaatccttgctgtgcgacagca.ttttataa;ttcaaaatttcagcagtaa 857
           ||| ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     862 GTTGTTCTGAAATTGATTTACCCAAACTACCAATCAATAAACTTTGAGAACCTGCACCCA 803

Qy     858 aattaaaatatgt 869
           |:|:|:|:|:|:|
Db     802 AATTAAAAGTCT 791


RESULT  10
US-08-680-726A-51/c
; Sequence 51, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-11
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  51:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10592 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-726A-51


  Query Match             3.4%;  Score 32.8;  DB 2;  Length 10592;
  Best Local Similarity   53.0%;  Pred. No. 2.6;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaaggaatttgaaatcattcatcattgttcatccatcttg 797
           |:|  ||| ||||   | |  || || ||   | | ||||| | | ||  |  | |  |
Db    4610 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 4551

Qy     798 gttcatcagaacaatccttgctgtgacacgaccttttataagttcaaaattcagcagtaa 857
           |||  || |||      || |   ||  |   |      || ||  | |  | ||||  |
Db    4550 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 4491

Qy     858 aattaaatatgt 869
           ||||||  | |
Db    4490 AATTAAAAGTCT 4479


RESULT  11
US-08-680-726A-52
; Sequence 52, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10592 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-726A-52


  Query Match             3.4%;  Score 32.8;  DB 2;  Length 10592;
  Best Local Similarity   53.0%;  Pred. No. 2.6;
  Matches   70;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     738 gatgattgacagacggttgaggaaggaatttgaaatcattcatcattgttcatccatcttg 797
           |:|  ||| ||||   | |  || || ||   | | ||||| | | ||  |  | |  |
Db    5983 GAAGGGTGATAGACAAATTATTAAAAAGTTCATAACTTTCATTAATTTTAAATGACCACG 6042

Qy     798 gttcatcagaacaatccttgctgtgacacgaccttttataagttcaaaattcagcagtaa 857
           |||  || |||      || |   ||  |   |      || ||  | |  | ||||  |
Db    6043 GTTGTTCTGAAATTGATTTACCCAAACTAGCAATCAATAAACTTTGAGAACCTGCAGCCA 6102

Qy     858 aattaaatatgt 869
           ||||||  | |
Db    6103 AATTAAAAGTCT 6114


RESULT  12
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D , Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  HCR96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-967-2


  Query Match             3.3%;  Score 31.8;  DB 2;  Length 2277;
  Best Local Similarity   32.2%;  Pred. No. 2.1;
  Matches   77;  Conservative   33;  Mismatches  129;  Indels    1;  Gaps    0;

Qy     244 gatgataacttggacagtccagatgaaattgaaatcaatgttatgaatttttatg 303
           |  |   |:| ||      | | ||| |   |:|   |:|   |   |     |  :
Db     553 GCNGTNGAYTGGGCNGTNGCNAARGAYAARTAYAARTGYAYNTAYGWNGTNWSNTTYTG 612

Qy     304 gatgaagcagattcttttgagtacactggtcatgat 339
           |  |:|  : | ::   :|||:  |   :|| ||  :       |  ||  :
Db     613 GCNGARGARAARWSNCAYGARWSNAARTAYGARGARWSNGTNAARAARAAYGCNWSNTC 672

Qy     364 caagagtcagagtctattccagaatatacatgcgaagagtgtttgtttatgtaa 417
           |:|||:    |      :||:  :   | :|| ||  :       :  :  :
Db     673 GARGARGAYATGGARGARGARAAYTAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAY 732

Qy     424 tggatgacagtggtcattggagaa                              482
```

```
Db    733 GGNGTNTTYGAYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCC 791


RESULT 13
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,974
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-974-2


  Query Match             3.3%;  Score 31.8;  DB 2;  Length 2277;
  Best Local Similarity   32.2%;  Pred. No. 2.1;
  Matches   77;  Conservative   33;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    244 gatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgatacccccc 303
          |  |   |:  ||     :  | || |   :|   |:     |     |    |   :
Db    553 GGNGINGAYIGGGCNGINGCNAARGAYAARIAYAARGAYACNCARWSNGTNWSNGCNATH 612

Qy    304 gatgaagcagattcttttgagtacactggccatgatcccacagccaacaaagattctggc 363
          |:|: : | ::   : | |:     :    :     :                    :
Db    613 GGNGARGARAARWSNCAYGARWSNAARGAYGARGARWSNGTNAARAARAARGGNMGNGAR 672

Qy    364 caaggagtcagagttctattccagaatatacggccgaagaggaacgggaggacaaccggctt 423
          | :|:  :  |     :|| |:  | :|| || ||    || ||: |   :  :
Db    673 GARGARGAYAIGGARGARGARGARAAYGAYCARGAYGAYGAYGAYGAYGAYGARGARGAY 732

Qy    424 tggatgacagtggtcattggagaacaagagcagcgcattgacatgaaggtcatcgagcc 482
          | |   | | :   |  || |:||: |:  :||:|| :  ||:|| |   |:||
Db    733 GGNGTNTTYGAYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCC 791


RESULT 14
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, Kathleen
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 42
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,487
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-098-487-2


  Query Match             3.3%;  Score 31.8;  DB 4;  Length 2277;
  Best Local Similarity   32.2%;  Pred. No. 2.1;
  Matches   77;  Conservative   33;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    244 gatgataacttggacagtccagatgaaattgacatcaatgtggatgaacttgatacccccc 303
          |  |   |:  ||     :  | || |   :|   |:     |     |    |   :
Db    553 GGNGTNGAYTGGGCNGTNTCNAARGAYAARATHAAYGAYACNCARWSNGTNWSNGCNATH 612

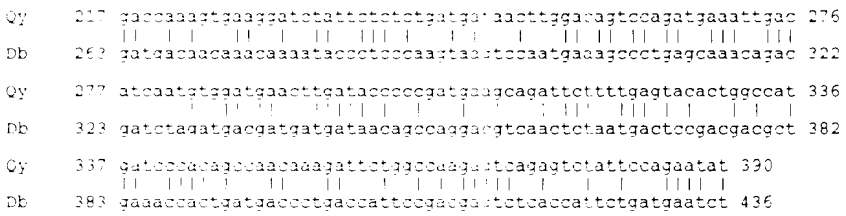
Qy    304 gatgaagcagattcttttgagtacactggccatgatcccacagccaacaaagattctggc 363
          |:|: : | ::   : | |:     :    :     :                    :
Db    613 GGNGARGARAARWSNCAYGARWSNAARGAYGARGARWSNGTNAARAARAARGGNMGNGAR 672

Qy    364 caaggagtcagagttctattccagaatatacggccgaagaggaacgggaggacaaccggctt 423
          | :|:  :  |     :|| |:  | :|| || ||    || ||: |   :  :
Db    673 GARGARGAYAIGGARGARGARGARAAYGAYCARGAYGAYGAYGAYGAYGAYGARGARGAY 732

Qy    424 tggatgacagtggtcattggagaacaagagcagcgcattgacatgaaggtcatcgagcc 482
          | |   | | :   |  || |:||: |:  :||:|| :  ||:|| |   |:||
Db    733 GGNGTNTTYGAYGAYGARGAYGARGARGARGARAAYATHGARWSNAARGTNACNAARCC 791


RESULT 15
5340934-7
;Patent No. 5340934
;    APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.;
, ROBEY, PAMELA G.
;    TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MORPHOGENETEINS
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/434,144
;      FILING DATE: 03-NOV-1989
;SEQ ID NO:7:
;      LENGTH: 1357
5340934-7


  Query Match             3.3%;  Score 31.6;  DB 4;  Length 1357;
  Best Local Similarity   48.9%;  Pred. No. 1.8;
  Matches   85;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;
```



Search completed: February 10, 2000, 02:58:51
Job time: 1509 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 10, 2000, 02:56:00 , Search time 159.56 Seconds
(without alignments)
808.011 Million cell updates/sec

Title: US-09-174-937-3
Perfect score: 1113
Sequence: 1 atggggaccaccgaagccac ... atcaggaaacaagcatgtcc 1113

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS4_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	231.6	20.8	1386	3	US-08-408-095-22	Sequence 22, Appl
2	52	4.7	3624	1	US-07-951-715A-6	Sequence 6, Appli
3	52	4.7	3624	3	US-08-459-448A-6	Sequence 6, Appli
4	43.6	3.9	3044	1	US-07-972-644-44	Sequence 44, Appl
5	43.6	3.9	3044	1	US-08-297-494-44	Sequence 44, Appl
6	43.6	3.9	3044	1	US-08-297-510-44	Sequence 44, Appl
7	43.6	3.9	3044	2	US-08-479-532-44	Sequence 44, Appl
8	43.6	3.9	3044	2	US-08-455-526-44	Sequence 44, Appl
9	43.6	3.9	3044	2	US-08-455-525-44	Sequence 44, Appl
10	43.6	3.9	3044	5	PCT-US92-03222-44	Sequence 44, Appl
c 11	43	3.9	1931	4	US-09-130-114-2	Sequence 2, Appli
c 12	42.4	3.8	30001	1	US-08-125-468-1	Sequence 1, Appli
c 13	42.4	3.8	30001	3	US-08-474-933-1	Sequence 1, Appli
14	41.8	3.8	4692	3	US-08-916-917-1	Sequence 1, Appli
15	41.8	3.8	4692	3	US-08-972-631-1	Sequence 1, Appli
16	41.8	3.8	4692	3	US-08-972-629-1	Sequence 1, Appli
17	41.8	3.8	4692	3	US-08-972-630-1	Sequence 1, Appli
18	41.8	3.8	4692	3	US-08-672-211-1	Sequence 1, Appli
19	41.8	3.8	1155	4	US-08-387-942C-7	Sequence 7, Appli
20	41.2	3.7	1485	1	US-08-471-601-23	Sequence 23, Appl
21	41.2	3.7	1485	1	US-08-471-556-23	Sequence 23, Appl
22	41.2	3.7	1485	2	US-08-479-382-23	Sequence 23, Appl
23	41.2	3.7	1485	2	US-08-479-354-23	Sequence 23, Appl
24	41.2	3.7	1485	2	US-08-479-383-23	Sequence 23, Appl
25	41.2	3.7	1485	3	US-08-479-041-23	Sequence 23, Appl
26	41.2	3.7	4897	6	5196516-7	Patent No. 5196516
c 27	40.2	3.6	322007	3	US-09-370-379-20	Sequence 20, Appl
28	40.2	3.6	3489	4	US-08-728-323A-1	Sequence 1, Appli
29	40	3.6	5690	3	US-08-447-464-2	Sequence 2, Appli
30	40	3.6	5690	3	US-08-716-679-2	Sequence 2, Appli
31	40	3.6	12588	4	US-09-387-942C-1	Sequence 1, Appli
32	40	3.6	1176	4	US-08-387-942C-17	Sequence 17, Appl
33	40	3.6	1143	4	US-08-387-942C-19	Sequence 19, Appl
c 34	39.4	3.5	2481	2	US-08-321-248-38	Sequence 38, Appl
c 35	39.4	3.5	2481	2	US-08-532-391-35	Sequence 35, Appl
36	39.4	3.5	32207	3	US-08-377-837-20	Sequence 20, Appl
c 37	39.4	3.5	3489	4	US-08-728-323A-1	Sequence 1, Appli
c 38	39.4	3.5	2481	5	PCT-US95-11511-35	Sequence 35, Appl
39	39.2	3.5	4039	2	US-08-363-301-1	Sequence 1, Appli
40	38.6	3.5	2277	2	US-08-670-997-5	Sequence 5, Appli
41	38.6	3.5	2277	2	US-08-674-974-5	Sequence 5, Appli
42	38.6	3.5	2277	4	US-09-199-487-5	Sequence 5, Appli
c 43	38.2	3.4	7218	1	US-08-232-465-14	Sequence 14, Appl
44	38.2	3.4	35091	3	US-08-752-790A-1	Sequence 1, Appli
45	37.6	3.4	5962	6	5386025-5	Patent No. 5386025

ALIGNMENTS

RESULT 1
US-08-408-095-22
; Sequence 22, Application US/08408095
; Patent No. 5858678
; GENERAL INFORMATION:
; APPLICANT: Chinnadurai, Govindaswamy
; TITLE OF INVENTION: APOPTOSIS-PROMOTING PROTEINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIGHPTE, MION, TINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,095
; FILING DATE: 21-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mack, Susan J.
; REGISTRATION NUMBER: 30,951
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 214..1155
US-08-408-095-22

Query Match 20.8%; Score 231.6; DB 3; Length 1386;
Best Local Similarity 56.8%; Pred. No. 3.3e-52;
Matches 551; Conservative 0; Mismatches 374; Indels 40; Gaps 5;

Qy 9 caccgaaggccacgctccggatgggggcgtgcgtgggtgggtgtgtgtgcgggaaa 68
Db 195 CATTGAATCCAGGCTGAGGA 254

Qy 69 tcttcccaggccactcccagaagaaacgggcgtggaagtgcttcgcaccccgtcgaaaa 128
Db 255 TTTTCCGATACCTTTACCAGAA 314

Qy 129 cacatcctctcctcctaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 188

```
Db    315 ACCAGAGGACCAGCCTGGCTCACTAGAAGTTAATGGAA---ATAAAGTGAGAAAGAAACT 371

Qy    189 agtggccccagagatcaacatttctctgaatcagagtgaggggtccctgctgtccgatga 248

Db    372 AATGGCTCCAGACATTAGCCTGACACTGGATTTTAGTGATGGCTCTGTATTGTCAGATGA 431

Qy    249 cttcttgaatacccctgatgacctggatattaacgtggatgacatcgagacccccgatga 308

Db    432 TTT------GGATGAAATTTTTGAGATTTTCTTAGATGCTTAGACACACCGTCAGA 482

Qy    309 gacggactggctggagttcctgggaattgccaacgaactggagtgggaagacgacacccc 368

Db    483 ----------------------GAATAGTAATGAGTTTGAGTGGGAAGATGATCTTCC 518

Qy    369 cgtggccaccgccaagaacatgccccgggacaagcgcggatctatttggggacggcacgac 428

Db    519 AAAACCCAAGACTACTGAAGTAATTAGGAAAGCTC------AATTACTGAATACACAGC 572

Qy    429 ggaggacggcagcgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaaga 488

Db    573 AGCAGAGGAAAAAGAAG---ATGGAGGACGCCTGGGTATGTTCAGGATTGGAGAACAGGA 629

Qy    489 gcaccgtatagacctgcacatgatcttgacatgaaagtggtcacccacggagggtta 548

Db    630 CCACAGGGTTGATATGAAGGCAATTGAACCCTATAAAAAAGTTATCAGCCATGGGGGATA 689

Qy    549 ctacggcggaaggcctcaacgccatcatcttcttcggcagcctacttccttccagacagcag 608

Db    690 TTATGGGGATGGATTAAATGCCATTGTTGTATTTGCTGTCTGTTTCATGCCTGAAAGTAG 749

Qy    609 cctccccgactaccactacatcatggaagaactacttcctgtacgtcatcagcagcttaga 668

Db    750 TCAGTTAATTATAGATACTGATGGACAATTTTTTAAATATGTTATTGGCACTTTGGA 809

Qy    669 gatcatgatgagctgaggctacatgatgctgtttcctgaacggtgccacgcccccggcggag 728

Db    810 GCTATTAGTAGCAGAAAACTACATGATAGTTTATTTAAATGGTGCAACAACTCGAAGAAA 869

Qy    729 gatgcctggaatcggctggctgaagaagtgctaccagatgatcgaccggaggttgcggaa 788

Db    870 AATGCCCAGTCTGGGATGGCTCAGGAAATGTTATCAGCAAATTGATAGAAGGTTACGGAA 929

Qy    789 agacctgaagtccttgatcatcgtccacccctcgtggttcattcggactgtgctggccat 848

Db    930 AAATCTAAAATCCCTAATCATTGTACATCCTTCTTGGTTTATCAGAACACTTCTGGCTGT 989

Qy    849 ctctgccctttcatcaacgtcaagttcatcaacaagatccagtacgtgcacagcttgga 908

Db    990 TACAAGACCATTTATTAGCTCGAAATTCAGTTAAAAAATTACATACGTGTTTAATTTGGC 1049

Qy    909 agacctggagcaactcatcctatggaacatttcagatccagactgcgtcctgcaata 968

Db   1050 AGAACTAGCAGAACTTGTCCCCATGGAATACTTGGCATACCAGAATGCATAAAACAAGT 1109

Qy    969 cgaaagaggaa 979

Db   1110 TGATCAAGAA 1119


RESULT 2
US-07-951-715A-6
; Sequence 6, Application US/07951715A
; Patent No. 5625136
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/951,715A
;      FILING DATE:  25-SEP-1992
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3624 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3621
;      OTHER INFORMATION:  /product= "Full-length, maize
;      OTHER INFORMATION:  optmized cryIB"
;      OTHER INFORMATION:  /note= "Disclosed in Figure 6."
US-07-951-715A-6


  Query Match             4.7%;  Score 52;  DB 1;  Length 3624;
  Best Local Similarity   45.7%;  Pred. No. 0.00011;
  Matches  181;  Conservative    0;  Mismatches  215;  Indels    0;  Gaps    0;

Qy    268 gacctggatattaacgtggatgacatcgagacccccgatgatgacgacctctttgattt 327

Db   1249 GGCCTGCAGCTGAAGGACAGCGAGACCGACGTGCCCCCCGACGACCACGAGTTTAA 1308

Qy    328 ctggggaatggcaacgaactggagtttgaggacgagaccccgtggccaccgccaagaac 387

Db   1309 TACGAGAGCTACAGCCACCGCCTGAGCCACATCGGCATCATCCTGCAGAGTCTGGTGAAT 1368

Qy    388 atgccccgggacagcgcggatctatttggggacggcacgacggacgacgtgcccgccgac 447

Db   1369 GTGCCCGTGTACAGCTGGACCCACCGCAGCGCCGACGGCACCAACACCATCGCCTTAA 1428

Qy    448 aacgggcgcctgtggcggacagtgatcatcggggagcaaagaagcaccgtatagacctgcac 507
```

```
Db    1429 CGCATCACCCAGATCCCCATGGTGAAGGCCAAGGAGTCGCTCAGGGCACCACCGTGGTG 1488

Qy     509 atgatccgggcttacatgaaagtggtcgccgcgggaggtactacggcgaaggcctcaac 567

Db    1489 CGCGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAACACCGGCGGCTTCGGC 1548

Qy     568 gccatcatcgtcttcgacctgttcttc gacagcgacctcccgaactacctactc 627

Db    1549 TTCATTTTCTTTACCGCCAACCCCGCCTACGGCTACCGCATCGGCTTCGGCATC 1608

Qy     628 atcatggagaacctcttcctgtacgtcatcaagagc 663

Db    1609 GCCAGCACCGTGGACTTCGACTTCTTCGTGAGCCGC 1644


RESULT   3
US-08-459-448A-6
; Sequence 6, Application US/08459448A
; Patent No. 5859336
;  GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION:  SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION:  INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ne  5859336arris Corporation
;      STREET:  Patent & Trademark Dept., 520 White Plains
;      STREET:  Rd., POB 2005
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,448A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/951,715
;      FILING DATE:  25-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/772,027
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40403
;      REFERENCE/DOCKET NUMBER:  DGC 1077/CIP/DIV4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8582
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3624 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3621
;      OTHER INFORMATION:  /gene= "cryIB"
;      OTHER INFORMATION:  /product= "Bacillus thuringiensis delta-
;      OTHER INFORMATION:  /note= "Disclosed in Figure 6."
US-08-459-448A-6


  Query Match             4.7%;  Score 52;  DB 9;  Length 3624;
  Best Local Similarity   45.7%;  Pred. No. 0.0011;
  Matches  101;  Conservative    0;  Mismatches  21;  Indels    0;  Gaps    0;

Qy     269 gacctggatattaacgtggatgtcatcaaaacagcgatccacgtcgtcccgtcatgaatt 327

Db    1249 GGCCTGCAGCTGAACCGACAGCCGACGCCAGCTTCCGTCGCCTGCAGGTCAGCGACGTT 1308

Qy     328 ctgagcaatgagcgaacgagcgtacagcgagtcgagcgaggagcgcaacaagctgaagctg 387

Db    1309 TACGAGAGCTACAGCTACCGCCTGAGCGTAGCGTCCTTCGCGAGCTCCGGCAGCAAC 1368

Qy     388 atcccgggcgacagcgtgcgagtgctttcgtgcgagcgcgcgtcgtcgcgcgcgcgcc 447

Db    1369 GTGCCCGTGTACAGCTGGACCCACCGCAGCGCCGACGTCACCAACACCATCGACCCC 1428

Qy     448 aacaggcgcgcctatggcggcgacacctcctacgtgcgcagcgtcttcaacagcggcgag 507

Db    1429 CGCATCACCCAGATCCCCATGGTGAAGGCCAAGGAGTCGCTCAGGGCACCACCGTGGTG 1488

Qy     508 atgatccgggcttacatgaaagtggtcgccgcgggaggtactacggcgaaggcctcaac 567

Db    1489 CGCGGCCCCGGCTTCACCGGCGGCGACATCCTGCGCCGCACCAACACCGGCGGCTTCGGC 1548

Qy     568 gccatcatcgtcttcgacctgttcttcctcccgagcgacctcccgaactacctactc 627

Db    1549 TTCATTTTCTTTACCGCCAACCCCGCCTACGGCTACCGCATCGGCTTCGGCATC 1608

Qy     628 atcatggagaacctcttcctgtacgtcatcaagagc 663

Db    1609 GCCAGCACCGTGGACTTCGACTTCTTCGTGAGCCGC 1644


RESULT   4
US-07-872-644-44
; Sequence 44, Application US/07872644
; Patent No. 5389627
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterase
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.  Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,644
;      FILING DATE:  19920420
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,456
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5389527and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30902
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3044 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2834
US-07-872-644-44


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtcagggtcccctgctgtccgatgacttct 253
          ||| || | |   |||  | || |  | |  ||||          || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccccctgatgaccctggatattaagctgtatgacatcgagaccccccgatgagaccg 313
          ||     |  ||   | |||  |  ||| |  |    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAACGAGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actgctggaagttcctgggaaatggcaacgaactggagtgggaagacgacaccccccgtgg 373
          |  | ||||   |||| |||  |  | |  || ||  |  |    ||     |
Db    559 ATACCCTGGTCGCCCTGCGGAGGCTGCAGCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy    374 ccaccgccaagaacatgccccgggacagcgcgaatctatttgaggacgccgacgacggagg 433
          ||   |||    | |  |||  ||  |  |||    | | |||  | |||  |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACGGCGGAAGACCAGAAGGGCCGCGCCGTGT 678

Qy    434 acggcagcgcgccgccaacgggcgcctgtgcgacgacatgatcatcttcatcaacatcgac 493
          ||  |  |  || |  |  ||     | |  | |   ||   ||| |  |  |  |
Db    679 ACACCGACCGCGACCGCAAGATCGTCGAACTGTGCGGGGAACTGTACGACTTCATCTTC 738

Qy    494 gtatagacctgcacatgatccgcctttcgatgaagtggtgcccgacggagggtactacg 553
          |      | || ||  | |  | ||| |  |  ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800


RESULT   5
US-08-297-494-44
; Sequence 44, Application US/08297494
; Patent No. 5580771
;  GENERAL INFORMATION:
;    APPLICANT:  Beaucage, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/297,494
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,456
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5580771and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30902
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3044 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2834
US-08-297-494-44


  Query Match             3.9%;  Score 43.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy    194 ccccagagatcaacatttctctggatcagagtcagggtcccctgctgtccgatgacttct 253
          ||| || | |   |||  | || |  | |  ||||          || | | || | |||
Db    439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy    254 tggataccccctgatgaccctggatattaagctgtatgacatcgagaccccccgatgagaccg 313
          ||     |  ||   | |||  |  ||| |  |    | |  |    | |  ||||
Db    499 TGGTGCACTGTGGCCAGCTGAGTGATAACGAGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy    314 actgctggaagttcctgggaaatggcaacgaactggagtgggaagacgacaccccccgtgg 373
          |  | ||||   |||| |||  |  | |  || ||  |  |    ||     |
Db    559 ATACCCTGGTCGCCCTGCGGAGGCTGCAGCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy    374 ccaccgccaagaacatgccccgggacagcgcgaatctatttgaggacgccgacgacggagg 433
          ||   |||    | |  |||  ||  |  |||    | | |||  | |||  |
Db    619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACGGCGGAAGACCAGAAGGGCCGCGCCGTGT 678

Qy    434 acggcagcgcgccgccaacgggcgcctgtgcgacgacatgatcatcttcatcaacatcgac 493
          ||  |  |  || |  |  ||     | |  | |   ||   ||| |  |  |  |
Db    679 ACACCGACCGCGACCGCAAGATCGTCGAACTGTGCGGGGAACTGTACGACTTCATCTTC 738

Qy    494 gtatagacctgcacatgatccgcctttcgatgaagtggtgcccgacggagggtactacg 553
          |      | || ||  | |  | ||| |  |  ||  ||||  | |     || |
Db    739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy    554 gc 555
          ||
Db    799 GC 800
```

```
RESULT   6
US-08-297-510-44
; Sequence 44, Application US/08297510
; Patent No. 5602019
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/297,510
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5602019and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30412
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3044 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2834
US-08-297-510-44


  Query Match             3.9%;  Score 41.6;  DB 1;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy     194 ccccagagatcaacatttctctggatcagagtgaggggtccctgctgtccgatgacttct 253
Db     439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy     254 tggataccctgatgacctggatattaacgtggatgacatcgagaccccccgatgagaccg 313
Db     499 TGGTGCACTGTGGCCAGCTGAGTGATAATGAGGAATGGAGCTTGCAGGCGGTGGAGAAGC 558

Qy     314 atctgctggagttcctgggggatggcaatgctggagtggtaagagcgacacccccgtgg 373
Db     559 ATACCCTGGTCGCCTGCGGAGGGTGCAGGTCTGCAGCAGTGCGGGCCCAGGGAGGCTC 618

Qy     374 ccacgccaagaacatgcccgggggacaggcgcgatctatttgggacggcacgacggagg 433
Db     619 CCCGAGCCGTCCAGAACCCCCCGAGGGAGTCCGAAGACTATAAAGGTGTGTCTGTC 678

Qy     434 acggcagcgccgccgaacgggcgcgctctcgcgggacagtgatctgtcgtttaatcgaatgg 493
Db     679 ACACCGACCGCGACGGCAAGATCCTCAACTTTGCTGTGCAACTCTATGATCTTGATCTCT 738

Qy     494 gtatagacctgcacatgatccgcc tttatctaaatgtgt t  a tcaatttgtttttg 553
Db     739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTTCAGTATGATAATCCTCCGTTCTTT 798

Qy     554 gc 555
Db     799 GC 800


RESULT   7
US-08-479-532-44
; Sequence 44, Application US/08479532
; Patent No. 5776752
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,532
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/247,494
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5776752and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30412
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-575
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3044 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2834
US-08-479-532-44


  Query Match             3.9%;  Score 41.6;  DB 2;  Length 3044;
```

```
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy     194 ccccagagatcaaacatttctctggatcaaagtaggggtccctgctgtccgatgacttct 253
           ||| || | |   ||| | || |   | |  |||          || | | || | |||
Db     439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy     254 tggatacccctgatgacctggatattaacctgaatgacatcgagacccccgatgagacag 313
           |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     499 TGGTGCACTGTGGCCAGCTGAGTGATAAAGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy     314 actcgctggagatttcctgggaaatggcaacgaaactggagtgggaagacgacacccccgtgg 373
           |  | |||    |||| |||  |   | |  ||| ||  |    |    ||     |
Db     559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy     374 ccaccgccaaagaacatgcccggggacagcgcggatctatttgaggacgacaacagcagc 433
           ||  |||    | |  ||| |||  |  |  |      | | |||  | |||  |
Db     619 CCCGAGCCGTCCAGAACCCCCCCGGAGGGGACGCCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy     434 acggcagcgccgccaacgggcgcctgtagcggacatgatcatcagcttcgtgaacgtgacc 493
           ||  |  |  || |  |  | |||    ||  |  |    ||  ||  |  |  |
Db     679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGGAACTCTACGACCTGGATGCCT 738

Qy     494 gtatagacctgcacatgatccggccttacatgaaagtggtccacccacggaggggtactacg 553
           |    | || ||  | |  |  ||| || |    ||  ||||  | |  |  |  |  |
Db     739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy     554 gc 555
           ||
Db     799 GC 800


RESULT   8
US-08-455-526-44
; Sequence 44, Application US/08455526
; Patent No. 5789553
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,526
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/297,194
;      FILING DATE:  29-AUG-1994
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5789553and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27900/30142
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3044 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2834
US-08-455-526-44


  Query Match             3.9%;  Score 43.6;  DB 2;  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy     194 ccccagagatcaaacatttctctggatcaaagtaggggtccctgctgtccgatgacttct 253
           ||| || | |   ||| | || |   | |  |||          || | | || | |||
Db     439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy     254 tggatacccctgatgacctggatattaacctgaatgacatcgagacccccgatgagacag 313
           |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     499 TGGTGCACTGTGGCCAGCTGAGTGATAAAGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy     314 actcgctggagatttcctgggaaatggcaacgaaactggagtgggaagacgacacccccgtgg 373
           |  | |||    |||| |||  |   | |  ||| ||  |    |    ||     |
Db     559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy     374 ccaccgccaaagaacatgcccggggacagcgcggatctatttgaggacgacaacagcagc 433
           ||  |||    | |  ||| |||  |  |  |      | | |||  | |||  |
Db     619 CCCGAGCCGTCCAGAACCCCCCCGGAGGGGACGCCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy     434 acggcagcgccgccaacgggcgcctgtagcggacatgatcatcagcttcgtgaacgtgacc 493
           ||  |  |  || |  |  | |||    ||  |  |    ||  ||  |  |  |
Db     679 ACACCGACCGCGACCGCAAGATCCTCCAACTGTGCGGGGGAACTCTACGACCTGGATGCCT 738

Qy     494 gtatagacctgcacatgatccggccttacatgaaagtggtccacccacggaggggtactacg 553
           |    | || ||  | |  |  ||| || |    ||  ||||  | |  |  |  |  |
Db     739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy     554 gc 555
           ||
Db     799 GC 800


RESULT   9
US-08-455-525-44
; Sequence 44, Application US/08455525
; Patent No. 5800987
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/455,525
;       FILING DATE:  31-MAY-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/297,494
;       FILING DATE:
;       APPLICATION NUMBER:  US 07/688,356
;       FILING DATE:  04-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No  5800987and, Greta E.
;       REGISTRATION NUMBER:  35,302
;       REFERENCE/DOCKET NUMBER:  27866/20822
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 346-5750
;       TELEFAX:  (312) 984-9740
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  44:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3044 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2834
US-08-455-525-44


  Query Match             3.9%;  Score 43.6;  DB 2:  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy     194 ccccagagatcaacatttctctggatcagagtgaggggtcctgctgtccgatgacttct 253
           ||| | | |   |||  | || |   |   ||||          || | | || | |||
Db     439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy     254 tggataccctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccg 313
           |||   |  ||   | |||  |  ||| | |  | | |   | |  | || ||||
Db     499 TGGTGCACTGTGGCCAGCTGAGTGATAAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy     314 actcgctggagttcctggggaatggcaacgaactggagtgggaagacgacaccccgtgg 373
           |  | ||||   |||| ||| |  | | .|| ||  |   |    ||     |
Db     559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy     374 ccaccgccaagaacatgccccggggacagccccgatctatttggggacggcacgacggagg 433
           ||   |||   | |  |||  |||  |  ||||     |     | | |||  | |||  |
Db     619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACCCCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy     434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaagagcacc 493
           ||  |  |  || |  |  |  |||    |||  |  |  |  ||  || |  ||   |
Db     679 ACACCGACCGCGACCGCAAGATCCTCCAACTCTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy     494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
           |     | || ||   | |  |  ||| | |   ||  |||| | |     || |
Db     739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy     554 gc 555
           ||
Db     799 GC 800


RESULT  10
PCT-US92-03222-44
; Sequence 44, Application PC/TUS9203222
;  GENERAL INFORMATION:
;     APPLICANT:  Beavo, Joseph A.
;     APPLICANT:  Bentley, Kelley
;     APPLICANT:  Charbonneau, Harry
;     APPLICANT:  Sonnenburg, William K.
;     TITLE OF INVENTION:  DNA Encoding Mammalian
;     TITLE OF INVENTION:  Phosphodiesterases
;     NUMBER OF SEQUENCES:  58
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;       ADDRESSEE:  Bicknell
;       STREET:  Two First National Plaza, 20 South Clark
;       STREET:  Street
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60603
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.27
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US92/03222
;       FILING DATE:  19920420
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/688,356
;       FILING DATE:  04-APR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Noland, Greta E.
;       REGISTRATION NUMBER:  35,312
;       REFERENCE/DOCKET NUMBER:  27866/20822
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 346-5750
;       TELEFAX:  (312) 984-9740
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  44:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3044 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2834
PCT-US92-03222-44


  Query Match             3.9%;  Score 43.6;  DB 5:  Length 3044;
  Best Local Similarity   45.0%;  Pred. No. 0.016;
  Matches  163;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy     194 ccccagagatcaacatttctctggatcagagtgaggggtcctgctgtccgatgacttct 253
           ||| | | |   |||  | || |   |   ||||          || | | || | |||
Db     439 CCCAAGTGCTGGTCATGCCGCTAGCGGACAAGGAGGCTGGGGCCGTGGCAGCTGTCATCT 498

Qy     254 tggataccctgatgacctggatattaacgtggatgacatcgagaccccgatgagaccg 313
           |||   |  ||   | |||  |  ||| | |  | | |   | |  | || ||||
Db     499 TGGTGCACTGTGGCCAGCTGAGTGATAAATGAGGAATGGAGCCTGCAGGCGGTGGAGAAGC 558

Qy     314 actcgctggagttcctggggaatggcaacgaactggagtgggaagacgacaccccgtgg 373
           |  | ||||   |||| ||| |  | | .|| ||  |   |    ||     |
Db     559 ATACCCTGGTCGCCCTGCGGAGGGTGCAGGTCCTGCAGCAGCGCGGGCCCAGGGAGGCTC 618

Qy     374 ccaccgccaagaacatgccccggggacagccccgatctatttggggacggcacgacggagg 433
           ||   |||   | |  |||  |||  |  ||||     |     | | |||  | |||  |
Db     619 CCCGAGCCGTCCAGAACCCCCCGGAGGGGACCCCGGAAGACCAGAAGGGCGGGGCGGCGT 678

Qy     434 acggcagcgccgccaacgggcgcctgtggcggacagtgatcatcggggagcaagagcacc 493
           ||  |  |  || |  |  |  |||    |||  |  |  |  ||  || |  ||   |
Db     679 ACACCGACCGCGACCGCAAGATCCTCCAACTCTGCGGGGAACTCTACGACCTGGATGCCT 738

Qy     494 gtatagacctgcacatgatccggccttacatgaaagtggtcacccacggagggtactacg 553
           |     | || ||   | |  |  ||| | |   ||  |||| | |     || |
Db     739 CTTCCCTGCAGCTCAAAGTGCTCCAATACCTGCAGCAGGAGACCCGGGCATCCCGCTGCT 798

Qy     554 gc 555
```

```
Db    799 GC 800


RESULT  11
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0967/12903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2


  Query Match             3.9%;  Score 4[illegible];  DB 4;  Length 1931;
  Best Local Similarity   45.6%;  Pred. No. 0.019;
  Matches  151;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy    177 [illegible] 236
          || | ||  ||||| | |    | ||| |  |    ||     || |   ||
Db   1049 GATGGAGGAGCTGGGGGCGGAGGTGATGGAGGAGGTGGGGCCGGAGGTGACGGAGGAGCT 990

Qy    237 [illegible] 296
          | | | | |        ||   |  || ||  |||      | | ||| ||   ||
Db    989 GGGGCCGGAGGTGGAGGACGAGGAGGGGGAGGAGGAGCCGGAGGAGGAGCGAGGACGG 930

Qy    297 [illegible] 356
          |      ||  |||  |||      ||||  |  || | | |  |||    |||  ||||
Db    929 GGAGGACGGGGAGGACGAGGACGGGGAGGAGGAGGAGGAGCGGGGAGGACGGGGA 870

Qy    357 [illegible] 416
           |||||   |   | || |   |  || ||  |  ||||||   || |||      |
Db    869 GGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGA 810

Qy    417 [illegible] 476
          |||||   |   | ||| | |   | |.  : |   |   |     |   | |
Db    809 GGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGA 750

Qy    477 [illegible] 537
            |.|.|  : ||. || |    ||  | ||
Db    749 GGGGAGGACGAGGACGGGGAGGACGAGGAG 719


RESULT  12
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the Biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,468
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-125-468-1


  Query Match             3.8%;  Score 42.4;  DB 1;  Length [illegible];
  Best Local Similarity   44.6%;  Pred. No. 0.085;
  Matches  166;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy    360 [illegible] 419
          || || |    |||||| |||| ||    || || || | |   |   | |
Db   8338 CGGCATCGGACTGGCCAGGGCCTA[illegible] 8279

Qy    420 [illegible] 479
          || |   ||| | | |||  |||||  |  |  ||   |  |  |  |  |
Db   8278 CGCCGCACCGCGCAGGGCTGC[illegible] 8219

Qy    480 [illegible] 539
          || | | |  | |    ||| | |  |  |  |  |  |  |  |  |  |
Db   8219 CGTGGACGGCCTCGCGGCCGACCTCGCCTCGCGCCGAGCTGACGACGCTTTTTTTTT 8159

Qy    540 [illegible] 599
          | | | ||  |  ||  |  ||  ||  |  | | |  |  |  |  |  |
Db   8159 GGCGGTGGACCGCTTCGGCGCGATCGCGCTCGTGGTCAACAACGCGGAGTTTTTTTTT 8099

Qy    600 [illegible] 659
          | |  || |  | ||  |  | ||  |  |  |  |  |  |  |  |  |
Db   8099 TGGCGTCACGTCCAGCGTCAGCGAGGAGTCGCGAGCGACGTGATCGACATTAACGTTGAT 8039

Qy    660 [illegible] 719
          |||| |    | | |    |||| |  ||  | ||  |  |  |  |  |  |
Db   8038 CAGCGTCTTCCGCGTCACCCGCGAGGTCCTCACGACGGGCGGCATCGAGGTTTTCTTTTA 7979

Qy    720 ccggcggaggat 731
          | ||||||  ||
Db   7978 CGGGCGGATCAT 7967


RESULT  13
US-08-474-933-1/c
; Sequence 1, Application US/08474933
; Patent No. 5866410
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
```

```
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,933
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,469
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-933-1


  Query Match             3.8%;  Score 42.4;  DB 3;  Length 30001;
  Best Local Similarity   44.6%;  Pred. No. 0.085;
  Matches  166;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy     360 cgacacccccgtggccaccgccaagaacatgccccggggacagcgcggatctatttgggga 419
           || || |    |||||| |||| ||    || || ||  | |    |    | ||
Db    8338 CGGCATCGGACTGGCCAGCGGCGAGGGGCTGGCCGCCGCGGCTGAGCGTCTTCATCTG 8279

Qy     420 cggcacgacgggaggacggcagcgccgccaacgagcgcctgtggcggacagtgatcatcgg 479
           || |   || | | | ||| |||||  |        ||   |||    | | | | |
Db    8278 CGCCGCCAGCGCCAGGGCGTCGCCGAGACGCTCGCCGCCTGCGCCGCGAGGGCCTGGA 8219

Qy     480 ggatcaagcgcacgtatagaactgcacatgatccggccttacatgaaagtggtcaccca 539
           || | | |  | |      ||| | | |     | |      |  |   | |||
Db    8218 GGTGGACGGCCTCGCGGCCGACGTCCGCTCCGCGCCGGACGTGAGCAGGCTCGTCCGGGC 8159

Qy     540 cggagggtactacggcgaaggcctcaacgccatcatcgtcttcgcagcctgcttccttcc 599
           | | | ||  |  ||  |  ||  ||   |  ||  |  ||| |   ||  |
Db    8158 GGCGGTGGACCGCTTCGGCGCGATCGGCGTGGTCAACAACGCCGGACGCAACGGCGG 8099

Qy     600 agacagcagcctccccgactaccactacatcatggagaacctcttcctgtacgtcatcag 659
           | | || | | | ||  |  | | | |  |  | || || |  ||    |  | |||
Db    8098 TGGCGTCACCTCCACCCTCACCGACGAGCTGTGGCAGGACGTGATCGACACCAACCTCAC 8039

Qy     660 cagcttagagctcctggtggctgaggactacatgatcgtgtacctgaacggtgccacgcc 719
           |||| |    | | |       |||| |  |   | ||    | || | |  |||  ||
Db    8038 CAGGGTCTTCCGCGTCACCCGCGAGGTCCTCACCACCGGCGCCATGGAGGCGGCCGGGCA 7979

Qy     720 ccggcggaggat 731
           | ||||||  ||
Db    7978 CGGGCGGATCAT 7967


RESULT  14
US-08-916-917-1
; Sequence 1, Application US/08916917
; Patent No. 5856132
;  GENERAL INFORMATION:
;    APPLICANT:  Stephens, Len
;    APPLICANT:  Hawkins, Phillip Thomas
;    APPLICANT:  Braselmann, Sylvia
;    TITLE OF INVENTION:  G-BETA GAMMA REGULATED
;    TITLE OF INVENTION:  PHOSPHATIDYLINOSITOL-3' KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds, LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,917
;      FILING DATE:  15-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/672,211
;      FILING DATE:  27-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Abrams, Samuel B
;      REGISTRATION NUMBER:  30,605
;      REFERENCE/DOCKET NUMBER:  8549-[illegible]-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4092 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-916-917-1


  Query Match             3.8%;  Score 41.8;  DB 3;  Length 4092;
  Best Local Similarity   55.1%;  Pred. No. 0.087;
  Matches  102;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy     855 cccttttcatcagcgtcaaqttcattaacaaagatccagtacatcaacagcttaaagaaat 914
           | | | |    |  | | |
Db    1232 CCCCATCCCGTCGCCAGGTGC AGAGTCAGAGCCTGAAGACAGCAGAGGTTTAAAGCGT 1291

Qy     915 ggagcaactcatccctatggaaaaggtccagatcaccaacgacaacgagcaaatgaaaaga 974
           | | | | | | |    |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db    1292 GCAGGAAATCCTGCTCAAGGAGCAGGAGCTGCTCCAGCCAGACATCGTGGACGACGAGA 1351

Qy     975 ggaa-agactgaaggccaggagggagagcaaggtcaaaaaaaaaaaaaaaaaaaaaaaaa 1033
           |||  ||   ||| |   | ||
Db    1352 GGACGAGGACGAGGAGGACGAGGAAAGGCTCCTGCTGCCCGACATCGCCGCGACGAGAG 1411

Qy    1034 ggtct 1038
           |||||
Db    1412 GGACT 1416


RESULT  15
US-08-972-631-1
; Sequence 1, Application US/08972631
; Patent No. 5856133
;  GENERAL INFORMATION:
;    APPLICANT:  Stephens, Len
;    APPLICANT:  Hawkins, Phillip T.
;    TITLE OF INVENTION:  G-BETA-GAMMA REGULATED
```

```
;    TITLE OF INVENTION:  PHOSPHATIDYLINOSITOL-3' KINASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  2730 Sand Hill Road
;      CITY:  Menlo Park
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,631
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/672,211
;      FILING DATE:  27-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P.
;      REGISTRATION NUMBER:  25,277
;      REFERENCE/DOCKET NUMBER:  8549-0005-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)854-3660
;      TELEFAX:  (415)854-3694
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4692 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-972-631-1


  Query Match             3.8%;  Score 41.8;  DB 3;  Length 4692;
  Best Local Similarity   55.1%;  Pred. No. 0.057;
  Matches  102;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy     855 ccctttcatcagcgtcaagttcatcaacaagatccagtacgtgcacagcttggaagacct 914
           | |  ||   | || || |  || | |  |  | ||  | ||||||| ||   |||
Db    1292 GGGCATCGGGGTCGCCAGGTGCTACGCTACAACCTGGAACGAGGACAGCTTCGACATCCT 1291

Qy     915 qgagcaactcatccctatggaacacgtccagatcccagactgcgtcctgcaatacgaaga 974
           | |  | || | |  | |||| |  |  |  ||          ||||| |  |||| ||
Db    1292 GCAGGAAATCCTGCTCAAGGAGCAGGAGCTGCTCCAGCCAGAGATCCTGGACGACGAGGA 1351

Qy     975 qqaa-agactgaaggccaggaggggaagcgcctggccccagccggagtttgtgctgccca 1033
           |||  ||   || |  | ||| ||||  |  || | | | |||    || || |
Db    1352 GGACGAGGACGAGGAGGACGAGGAAGAGGACTTGGACGCCGACGGCCACTGCGCGGAGAG 1411

Qy    1034 qgtct 1038
           | ||
Db    1412 GGACT 1416


Search completed: February 10, 2000, 02:57:04
Job time: 1402 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1999 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2000, 02:57:11 ; Search time 159.56 Seconds
(without alignments)
3074.508 Million cell updates/sec

Title: US-09-174-937-4
Perfect score: 4235
Sequence: 1 gtcgacccacgcgtccgcgg.....aaaaaaaaaagggcggccgc 4235

Scoring table: IDENTITY_NUC

Searched: 207703 seqs, 57918364 residues

Database : Issued_Patents_NA.*

Word size : 0

Number of hits that pass the threshold : 415406
1: /cgn2_6/ptodata/1/ina/TA_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/TP_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/TS_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/US_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/UTS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	286.2	6.8	1396	9	US-09-408-095-22	Sequence 22, Appl
2	80.6	1.9	7218	1	US-08-812,469-14	Sequence 14, Appl
c 3	63.8	1.5	5852	1	US-07-867,106-2	Sequence 2, Appli
4	59.8	1.4	2447	4	US-09-214,969-14	Sequence 14, Appl
5	58.9	1.4	1172	1	US-07-945,289-9	Sequence 9, Appli
6	58.8	1.4	1172	1	US-08-462,831-9	Sequence 9, Appli
7	58.8	1.4	1172	2	US-08-461,809-9	Sequence 9, Appli
8	58.8	1.4	1172	2	US-08-461,441-9	Sequence 9, Appli
9	58.8	1.4	1172	5	PCT-US93-08518-9	Sequence 9, Appli
10	58.6	1.4	240	1	US-08-628,417-6	Sequence 6, Appli
11	57	1.3	1558	1	US-08-455,550-7	Sequence 7, Appli
12	57	1.3	1461	5	PCT-US95-04258-4	Sequence 4, Appli
13	56.6	1.3	1641	1	US-08-301,903A-8	Sequence 8, Appli
14	55	1.3	2950	6	5058287[23]	Patent No. 5058287
15	54.4	1.3	1738	3	US-09-[illegible]482A-2	Sequence 2, Appli
16	54.2	1.3	19124	4	US-08-[illegible]826B-13	Sequence 13, Appl
17	53.8	1.3	1534	1	US-08-301,903A-6	Sequence 6, Appli
18	53.6	1.3	1066	2	US-08-[illegible]101A-4	Sequence 4, Appli
19	53.4	1.3	1493	1	US-09-[illegible]820-24	Sequence 24, Appl
20	53.4	1.3	1493	1	US-08-[illegible]535-24	Sequence 24, Appl
21	53.2	1.3	144	2	US-09-[illegible]344-26	Sequence 26, Appl
22	53.2	1.3	3239	5	PCT-US94-10080-5	Sequence 5, Appli
23	52.8	1.2	1338	3	US-08-[illegible]933B-7	Sequence 7, Appli
24	52.8	1.2	3591	3	US-09-[illegible]849-1	Sequence 1, Appli
25	52.8	1.2	1338	4	US-08-[illegible]158B-7	Sequence 7, Appli
26	52.8	1.2	1338	4	US-08-[illegible]596A-7	Sequence 7, Appli
27	52.8	1.2	1338	4	US-08-[illegible]150A-7	Sequence 7, Appli
28	52.2	1.2	2671	6	5168051-4	Patent No. 5168051
29	52	1.2	6671	1	US-08-[illegible]443-1	Sequence 1, Appli
30	52	1.2	2628	1	US-09-[illegible]219-1	Sequence 1, Appli
31	52	1.2	6671	1	US-09-[illegible]459-1	Sequence 1, Appli
32	52	1.2	6671	2	US-08-[illegible]678-1	Sequence 1, Appli
33	52	1.2	6671	5	PCT-US95-[illegible]275-1	Sequence 1, Appli
c 34	51.8	1.2	3001	2	US-08-[illegible]111-1	Sequence 1, Appli
c 35	51.8	1.2	2993	2	US-08-[illegible]10-1	Sequence 1, Appli
c 36	51.8	1.2	643	2	US-08-[illegible]10-17	Sequence 17, Appli
c 37	51.8	1.2	602	2	US-08-[illegible]10-18	Sequence 18, Appli
38	51.8	1.2	3001	2	US-08-[illegible]10-1	Sequence 1, Appli
39	51.8	1.2	2993	2	US-08-[illegible]111-1	Sequence 11, Appl
40	51.8	1.2	642	2	US-08-[illegible]10-[illegible]	Sequence 10, Appl
41	51.6	1.2	903	5	PCT-US95-[illegible]84-21	Sequence 21, Appl
42	51.4	1.2	972	1	US-08-[illegible]937-1	Sequence 1, Appli
43	51.4	1.2	972	1	US-08-[illegible]828-1	Sequence 1, Appli
44	51.4	1.2	1512	4	US-08-[illegible]-9650-8	Sequence 8, Appli
45	51.4	1.2	689	6	5149894-3	Patent No. 5149894

ALIGNMENTS

RESULT 1
US-08-408-095-22
; Sequence 22, Application US/08408095
; Patent No. 5858678
; GENERAL INFORMATION:
; APPLICANT: Chinnadurai, Govindaswamy
; TITLE OF INVENTION: APOPTOSIS-REGULATING PROTEINS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: [illegible] Main, 10th, McGuireWoods & Judge
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.[illegible] Version #1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/[illegible]
; FILING DATE: 21-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mack, Susan G.
; REGISTRATION NUMBER: 30,951
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 214..1155
US-08-408-095-22

Query Match 6.8%; Score 286.2; DB 9; Length 1396;
Best Local Similarity 61.7%; Pred. No. 2e-[illegible];
Matches 513; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

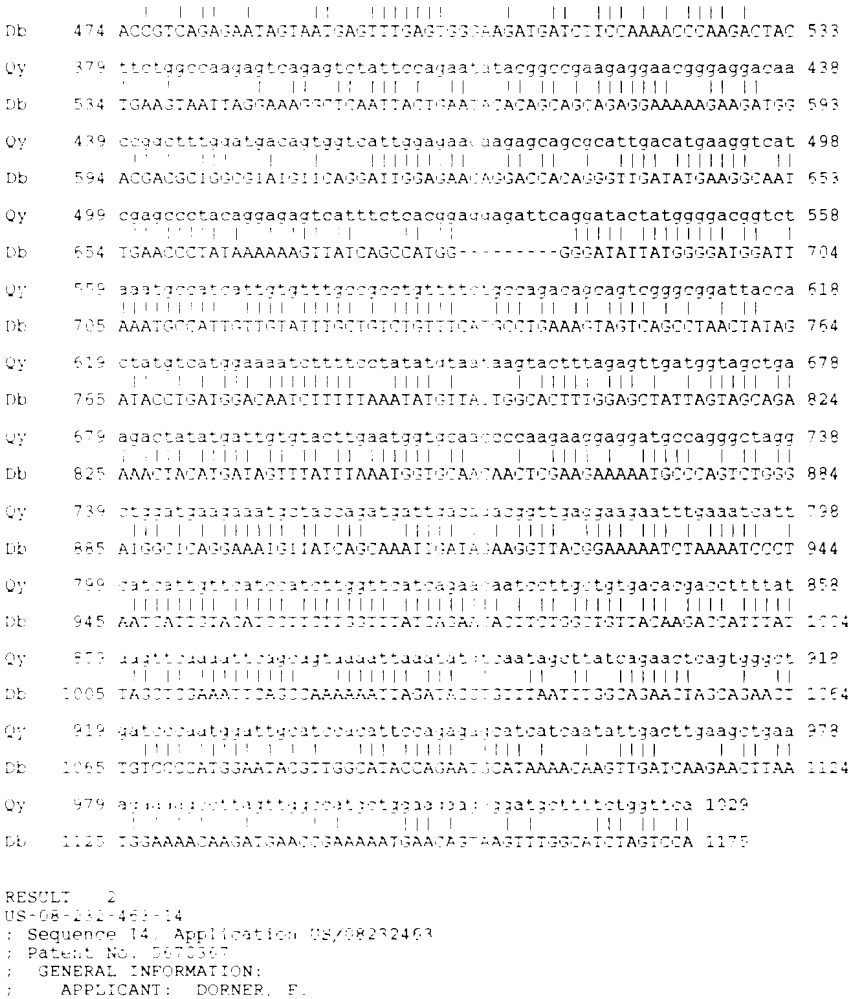
Qy 199 gagaatcaagctcaagagcccccaa[illegible] 258

Db 361 GAGAAAGAAACTAATGCCTCCAGA[illegible] 419

Qy 259 tctctctgatgataacttggacgat[illegible] 318

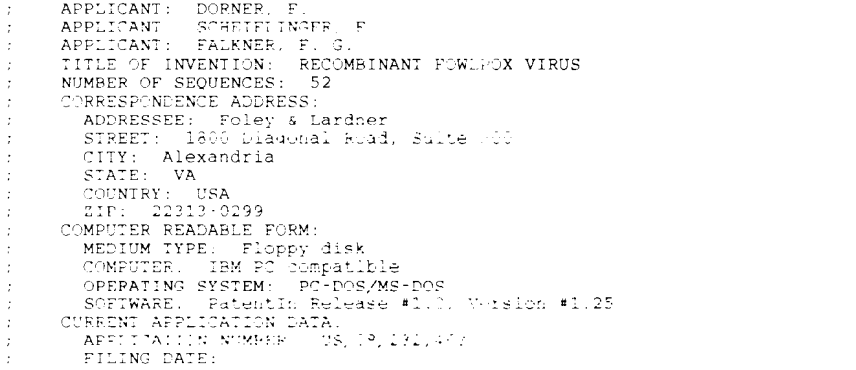
Db 420 ATTGTCAGATCATTTCGATCAAAGT---CCC[illegible] 473

Qy 319 taccccgatgaagcgagattcttt[illegible] 378



RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5672367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT SCHEIFLINGER, F
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US, 08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 1.9%; Score 80.8; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred. No. 5.6e-11;
Matches 25; Conservative 241; Mismatches 148; Indels 7; Gaps 0;



RESULT 3
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia

; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920605
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: PIPE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
US-07-867-106-2

Query Match 1.5%; Score 68.8; DB 1; Length 5852;
Best Local Similarity 53.4%; Pred. No. 1.1e-06;
Matches 134; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Qy 3976 ttagaaacatttagatgtgcacaaattaatcttttatatatctaaaggtttttctatcat 4035
Db 5841 TTACAAAGTAACAATAGTACTAATAAAGATCTTTTAAATTTATTAATATACATTTTTAAT 5782

Qy 4036 gcattggattgctcagaataaagtgtctgttagacttcgttttggtaaataaattctcca 4095
Db 5781 GGTTATTTAATTTATTTAATTATTTGTTATTTGTATTTTTTTATATATGTTATTTGTTGT 5722

Qy 4096 taatgtagattaatatatataaagtctttaatgacacaatatatctatatagcctcactg 4155
Db 5721 TTGTTTTTTTACTTATATTTCTATTTTTTATTTATAAAATTAATAATTAAATTTTAAATA 5662

Qy 4156 tataattcagaaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4215
Db 5661 AATAAAAAAAAAAAAAAAAAAAAAAATTTAAATTAAAAAAAAAAAAAAAAAAAAAAAAA 5602

Qy 4216 aaaaaaaaag 4226
Db 5601 AATAAAAAATG 5591

RESULT 4
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,923
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-014-969-14

Query Match 1.4%; Score 59.8; DB 4; Length 2447;
Best Local Similarity 52.7%; Pred. No. 7.4e-06;
Matches 127; Conservative 1; Mismatches 113; Indels 1; Gaps 0;

Qy 3985 tttagatgtgcacaaattaatcttttatatatctaaaggtttttctatcatcattggat 4044
Db 2161 TATAGATATGTATATATGGTTCACATATCTGGATCTGTGTATTTGATTTTGTATTTTAAA 2220

Qy 4045 tgctcagaataaagtgtctgttagacttcgttttggtaaataaattctccataatatata 4104
Db 2221 TGTGACAAATAAACCTTTTGGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2280

Qy 4105 ttaatatatataaaagtctttaatgacacaatatatctatatagcctcactatatagttcg 4164
Db 2281 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2340

Qy 4165 gaaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4224
Db 2341 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2400

Qy 4225 a 4225
Db 2401 A 2401

RESULT 5
US-07-945-288-9
; Sequence 9, Application US/07945288
; Patent No. 5433948
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne P.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: CLONING AND SEQUENCING [illegible]
; TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)

```
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 51
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,288
;      FILING DATE:  19920910
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  P36,207
;      REFERENCE/DOCKET NUMBER:  IPC-01000 (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-07-945-288-9

  Query Match             1.4%;  Score 56.8;  DB 1;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-06;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   4076 tttggtaaataaattctccataatgtagattaatatataaaagtctttaatgacacaat 4135
          ||  |||| |||   | ||    | |  |  |  |  |||   ||  |||  |
Db   1005 TTCTCTAAACAAAAAGACAATTTCTATATGATTGTCACTAATTTATTTAAAATCAAAAT 1064

Qy   4136 atatctatatagcctcactgtatattttcaaaaataaaaattgattctgcaaaaaaaaaaaa 4195
          | | | | |  | | |  ||     | |  ||||||   |      ||||||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCACAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          |||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT   6
US-08-462-831-9
; Sequence 9, Application US/08462831
; Patent No. 5952142
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 51
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,831
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
;      FILING DATE:  10 SEPTEMBER 1992
;      APPLICATION NUMBER:  US 580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  US 458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-01000 (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-08-462-831-9

  Query Match             1.4%;  Score 56.8;  DB 1;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-06;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   4076 tttggtaaataaattctccataatgtagattaatatataaaagtctttaatgacacaat 4135
          ||  |||| |||   | ||    | |  |  |  |  |||   ||  |||  |
Db   1005 TTCTCTAAACAAAAAGACAATTTCTATATGATTGTCACTAATTTATTTAAAATCAAAAT 1064

Qy   4136 atatctatatagcctcactgtatattttcaaaaataaaaattgattctgcaaaaaaaaaaaa 4195
          | | | | |  | | |  ||     | |  ||||||   |      ||||||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCACAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          |||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT   7
US-08-461-809-9
; Sequence 9, Application US/08461809
; Patent No. 5770202
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  ASCII TEXT
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/461,809
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/945,288
;       FILING DATE:  10 SEPTEMBER 1992
;       APPLICATION NUMBER:  US 580,655
;       FILING DATE:  11 SEPTEMBER 1990
;       APPLICATION NUMBER:  US 458,642
;       FILING DATE:  13 FEBRUARY 1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MANDRAGOURAS, AMY E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 227-7400
;       TELEFAX:  (617) 227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1172 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..738
US-08-461-809-9


  Query Match             1.4%;  Score 58.8;  DB 2;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-06;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   4076 tttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacacaat 4135
          ||  |||| |||   | ||   || ||  |   |  ||  | |||||   || |||
Db   1005 TTCTCTAAACAAAAAGACAATTTCTTATATGATTGTCACTAATTTATTTAAAATCAAAAT 1064

Qy   4136 atatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaaaa 4195
          | |  | | | | | ||    | | ||||||   |       ||||||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCACAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT   8
US-08-461-441-9
; Sequence 9, Application US/08461441
; Patent No. 5773002
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,441
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
;      FILING DATE:  10 SEPTEMBER 1992
;      APPLICATION NUMBER:  US 580,655
;      FILING DATE:  11 SEPTEMBER 1990
;      APPLICATION NUMBER:  US 458,642
;      FILING DATE:  13 FEBRUARY 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
US-08-461-441-9


  Query Match             1.4%;  Score 58.8;  DB 2;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-06;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   4076 tttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacacaat 4135
          ||  |||| |||   | ||   || ||  |   |  ||  | |||||   || |||
Db   1005 TTCTCTAAACAAAAAGACAATTTCTTATATGATTGTCACTAATTTATTTAAAATCAAAAT 1064

Qy   4136 atatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaaaa 4195
          | |  | | | | | ||    | | ||||||   |       ||||||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCACAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT   9
PCT-US93-08518-9
; Sequence 9, Application PC/TUS9308518
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
;    TITLE OF INVENTION:  DERMATOPHAGOIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII TEXT
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08518
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/945,288
```

```
;      FILING DATE:  10 SEPTEMBER 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MANDRAGOURAS, AMY E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-010CC (IMI-024)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1172 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..738
PCT-US93-08518-9


  Query Match             1.4%;  Score 58.8;  DB 5;  Length 1172;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-06;
  Matches   93;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   4076 tttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacacaat 4135
          |  |  ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |||
Db   1005 TTGTCTAAACAAAAAGACAAITTGTTATATGATGTCACTAATTTATTTAAAATCAAAAT 1064

Qy   4136 atatctatatagcctcactgtataattcagaaataaaaaattgattctgcaaaaaaaaaaaa 4195
          |  |  |  |  |  |  |  |  ||     |  ||||||   |     ||||||||||||
Db   1065 TTTTAGAAAATGAATAAATTCATTCACAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAA 1124

Qy   4196 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||||
Db   1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1154


RESULT  10
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
;  GENERAL INFORMATION:
;    APPLICANT:  GILLESPIE, DAVID
;    TITLE OF INVENTION:  COMPETITOR PRIMER ASYMMETRIC
;    TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  U.S. ARMY CHEMICAL AND BIOLOGICAL
;      ADDRESSEE:  DEFENSE COMMAND
;      STREET:  OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
;      CITY:  ABERDEEN PROVING GROUND
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21010-5423
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,417
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIFFONI, ULYSSES J
;      REGISTRATION NUMBER:  39,908
;      REFERENCE/DOCKET NUMBER:  DAM 398-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  410-671-1158
;      TELEFAX:  410-671-2534
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligodeoxynucleotide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  YES
US-08-628-417-6


  Query Match             1.4%;  Score 58.6;  DB 1;  Length 241;
  Best Local Similarity   60.2%;  Pred. No. 4.7e-06;
  Matches   97;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    0;

Qy   4065 ttagacttcgttttggtaaataaatttctccataatgtagattaataatataaaagtctttt 4124
          |||||  |  ||||  |||| |||      | ||  |   ||  | | | | | | | | |
Db     22 TTAGAAATAATTTTATTAAAAAAATATAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAA 82

Qy   4125 aatgacacaaatatatctatatagcctcactgtataattcagaaataaaaaattgattctgc 4184
          ||  | | ||  | |   | | |     :     ||     | | | | | | | | | | |
Db     83 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 142

Qy   4185 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||||
Db    143 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 183


RESULT  11
US-08-455-550-7
; Sequence 7, Application US/08455550
; Patent No. 5670338
;  GENERAL INFORMATION:
;    APPLICANT:  MURAKAMI, KAZUO
;    APPLICANT:  UENO, NAOTO
;    APPLICANT:  KATO, YUKIO
;    TITLE OF INVENTION:  XENOPUS INSULIN RING MORPHOGENETIC PE, DRIVE AND USE THE
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dike, Bronstein, Roberts & Cushman
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,550
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/056,564
;      FILING DATE:  30-APR-1993
;      APPLICATION NUMBER:  07/577,892
;      FILING DATE:  05-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eisenstein, Ronald I
;      REGISTRATION NUMBER:  30628
;      REFERENCE/DOCKET NUMBER:  40302-PWC-DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:  200291
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-455-550-7


  Query Match              1.3%;  Score 57;  DB 1;  Length 1558;
  Best Local Similarity   58.6%;  Pred. No. 3e-05;
  Matches   99;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    4053 ataaagtgtctgttagacttcgttttggtaaataaattctccataatgtagattaataat 4112
           || |  | | |||   | |   |||| |  |':|   ||        |  | |  |   ||
Db    1382 AIGACIIIIAIGIGIAAAIGIIIIIIIGACAAIAIGATCATATATTTTGACAAAATATAT 1441

Qy    4113 ataaaagtctttaatgacacaatatatctatatagcctcactgtataattcagaaataaa 4172
           || || |   || | | | || | |   . :'|   | | |   || || | | |||||
Db    1442 TTATAACIACGIAIIAAAAGAAAAAAAAAAAAAIAAAATAAGICATTATTTTAAACATAAA 1501

Qy    4173 aattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4221
           ||   |      |||||||||||||||||||| ||||||||||||||||
Db    1502 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1550


RESULT  12
PCT-US95-04258-4
; Sequence 4, Application PC/TUS9504258
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  A DNA MOLECULE ENCODING A MAST CELL
;    TITLE OF INVENTION:  FUNCTION-ASSOCIATED ANTIGEN (MAFA)
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04258
;      FILING DATE:  06-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  IL 109257
;      FILING DATE:  08-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  PECHT-1 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1461 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  54..617
PCT-US95-04258-4


  Query Match              1.3%;  Score 57;  DB 5;  Length 1461;
  Best Local Similarity   63.5%;  Pred. No. 2.9e-05;
  Matches   87;  Conservative    1;  Mismatches   5;  Indels    1;  Gaps    0;

Qy    4084 ataaatttctccataatgtagattaataatataaa'tatttaatcccccaatat'ctat 4143
           |' '  |  ||| ||||'  '  '  '  '  '  '  '  '  '  '  '  '
Db    1325 ATTATCTATCCCTAATTTCCCTAAAAATTAATAAAATTCCAATTAAAAATAAAAAATATTT 1384

Qy    4144 atagcctcactgtataattcagaaataaaaaattgattctgtaaaa.aaaaaaaaaaaaa 4203
           |   || |    | ||  |       .: |
Db    1385 TTCTTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1444

Qy    4204 aaaaaaaaaaaaaaaa 4220
           ||||||||||||||||
Db    1445 AAAAAAAAAAAAAAAA 1461


RESULT  13
US-08-300-903A-8
; Sequence 8, Application US/08300903A
; Patent No. 5591630
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, Dirk M
;    APPLICANT:  Giri, Judith G
;    TITLE OF INVENTION:  Interleukin-15 Receptors
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Operating System 7.1
;      SOFTWARE:  Microsoft Word for Apple, Version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/300,903A
;      FILING DATE:  06-SEPTEMBER-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 08/236,909
;      FILING DATE:  06-MAY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,694
;      REFERENCE/DOCKET NUMBER:  2822-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-587-0430
;      TELEFAX:  206-233-0644
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1641 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..839
US-08-300-903A-8


  Query Match              1.3%;  Score 56.6;  DB 1;  Length 1441;
  Best Local Similarity   56.8%;  Pred. No. 3.9e-05;
  Matches  104;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    0;

Qy    4043 attgctcagaaataaaatgtgtctgttagacttcgttttggtaaataaattctccataat 4102
           |||  | |  |  | ' ' | |
Db    1434 ATTTTACGAATATATGCGGTTTATATTTATACTTCCTTGTCTATTATATTCTATACTATATA 1493
```

```
Qy   4103 gattaatatataaaagtctttaatgacacaatatatctatatagcctcactgtataatt 4162
          |||| |  | |  |  |  | |  |  ||  | | ||  | |  |||
Db   1494 TATTATTTGTATTTTGACATTGTACCTTGTATAAACAAAATAAAACATCTATTTTCAATA 1553

Qy   4163 cagaaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4222
          |||| |||| |  |  |  |||||| || ||||||||||||||||||||||||||||||
Db   1554 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613

Qy   4223 aaa 4225
          |||
Db   1614 AAA 1616


RESULT  14
5258287-23
:Patent No. 5258287
;     APPLICANT: BAXTER, ROBERT C.;WOOD, WILLIAM I.
;     TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION
;OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53
;     NUMBER OF SEQUENCES: 58
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/171,713
;       FILING DATE: 22-MAR-1988
:SEQ ID NO:23:
;       LENGTH: 2550
5258287-23


  Query Match             1.3%;  Score 55;  DB 6;  Length 2550;
  Best Local Similarity   60.3%;  Pred. No. 0.00012;
  Matches   91;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy   4075 ttttggtaaataaatttctccataatgtagattaataatataaaagtctttaatgacacaa 4134
          |||| | | |   ||  |  |  |   |  |  || |   ||   ||  |    |||
Db   2381 ttttctttatttttttaacttttgggggaaaaagtatttttgagaagtttgtcttgcaa 2440

Qy   4135 tatatctatatagcctcactgtataattcagaaataaaaattgattctgcaaaaaaaaaa 4194
          | ||| |||| |   | | |  |   || |    ||||||   |      ||||||||||
Db   2441 tgtatttataaatagtaaatagagttttaccattaaaaaaaaaaaaaaaaaaaaaaaaaa 2500

Qy   4195 aaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||||||||||||||||||||||||||||| ||
Db   2501 aaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2531


RESULT  15
US-08-379-482A-2
; Sequence 2, Application US/08379482A
; Patent No. 5859334
;  GENERAL INFORMATION:
;     APPLICANT:  Brugliera, Filippa
;     APPLICANT:  Holton, Timothy A.
;     TITLE OF INVENTION:  GENETIC SEQUENCES ENCODING
;     TITLE OF INVENTION:  GLYCOSYLTRANSFERASE ENZYMES AND USES THEREFOR
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Scully, Scott, Murphy & Presser
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/379,482A
;       FILING DATE:  30-JUL-1993
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DiGiglio, Frank S.
;       REGISTRATION NUMBER:  31,346
;       REFERENCE/DOCKET NUMBER:  9590
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (516)742-4343
;       TELEFAX:  (516)742-4366
;       TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1738 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1413
US-08-379-482A-2


  Query Match             1.3%;  Score 54.4;  DB 3;  Length 1738;
  Best Local Similarity   65.8%;  Pred. No. 0.00015;
  Matches   75;  Conservative    0;  Mismatches   41;  Indels    ...  Gaps    0;

Qy   4106 taataatataaaagtctttaatgacacaatatatctatatagcctcactgtataattca 4165
          | || | || ||  |  |   |  || || || |  |||  |
Db   1614 TGATTAAATGAAGATGGTCTTTACCATAACATTTAAATAAACGATCATATTCATTTTAAA 1673

Qy   4166 aaataaaaattgattctgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 4225
          ||| |||||   |      ||||||||||||||||||||||||||||||||||||||||
Db   1674 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1733


Search completed: February 10, 2000, 02:58:19
Job time: 1477 sec
```

```
CM of: US-09-174-937-5 vs  Issued_Patents_NA *   Out_format : pfs

Date: Feb 7, 2... 5:16 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.mod3el -DEV-x1p
 -Q=/cgn2_1/USPTO_sp001/US09174937/runat_10022000_095108_13295/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -UGAPOP=4.500 -UGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-174-937-5
Query length: 322
Database: Issued_Patents_NA:*
Database sequences: 227703
Database length: 57918364
Search time (sec): 58.270000



SCuts_list:
Sequence          Strd Orig    ZScore   PScore Len  ! Documentation
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US98-22 -   862.50 1977.15  1.0e-101  1996  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-474-134-5 -    90.50  171.97  6.0e-42   1236  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-577-857A-7 -   88.00  155.99  0.2884    2722  !
/cgn2_6/ptodata/1/ina/5E_COMB.seq:US-08-666-271-1 -    87.50  153.19  0.6067    3906  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-711-340-1 -    87.00  137.48  3.10      9606  !
/cgn2_6/ptodata/1/ina/FA_COMB.seq:US-08-284-548A-1 -   87.00  137.48  3.10      9606  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-480-654-1 -    87.00  137.48  3.10      9606  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-177-235A-1 -   87.00  137.48  3.10      9606  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-[illegible]-326-29 -  86.00  147.39  0.8741  3706  !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5178[illegible]-8 -   84.50  156.09  0.1962  1137  ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-[illegible]-475A-4 -  82.00  145.94  1.76  2111  !
/cgn2_6/ptodata/1/ina/FB_COMB.seq:US-08-482-655B-1 -   82.00  127.30  11.43  9530  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-408-249A-4 -   82.00  145.94  1.76  2111  !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5248[illegible]-3 -  82.00  138.13  2.85  3666  ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-[illegible]-611-1 -  81.00  134.11  4.78  4181  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-874-1 -  81.00  134.11  4.78  4181  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-408-186-1 -   81.00  134.11  4.78  4181  !
/cgn2_6/ptodata/1/ina/FB_COMB.seq:US-08-482-549-1 -   81.00  134.11  4.78  4181  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-408-850-1 -   81.00  134.11  4.78  4181  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-[illegible]-792-2 -  81.00  129.82  8.41  8315  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-610-1 -  80.50  135.83  3.83  3937  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-837-6 -  80.00  156.08  0.2853  688  ! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-719-822B-1 -   80.00  135.83  3.83  3937  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US98-11310-9 -  80.50  156.08  0.2853  688  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-742-15 -  79.50  129.69  8.41  4486  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-[illegible]-913-1 -  79.00  116.75  44.26  11226  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US98-01034-1 -  79.50  140.54  2.09  1604
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-742-12 -  78.00  126.64  12.44  4328  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-446-[illegible]55A-1 -  78.00  117.37  40.87  9920  !
/cgn2_6/ptodata/1/ina/FD_COMB.seq:US-08-[illegible]-793-1 -  78.00  121.79  23.25  5328  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-300-1 -  77.50  126.23  13.12  4080  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-[illegible]-273-4 -  77.50  134.01  4.83  2223  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-06430-20 -  77.50  145.16  1.76  932
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-16993-1 -  77.50  126.22  19.11  4182
/cgn2_6/ptodata/1/ina/backfiles1.seq:5225[illegible]-4 -  77.50  137.81  2.87  1674  ! Pate
/cgn2_6/ptodata/1/ina/backfiles1.seq:52[illegible]-1 -  77.50  129.77  8.32  3195  ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-089-1 -  77.00  121.76  23.25  5275  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-565-1 -  77.00  121.76  23.25  5275  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-15 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-17 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-19 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-22 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-23 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-24 -  77.00  122.20  22.00  5100  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-[illegible]-292B-25 -  77.00  122.20  22.00  5100  !

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-408-095-22

seq_documentation_block:
; Sequence 22, Application US/08/408095
; Patent No. 5858678
; GENERAL INFORMATION:
;   APPLICANT:  Chinnadurai, Govindasamy
;   TITLE OF INVENTION:  APOPTOSIS-REGULATING PROTEINS
;   NUMBER OF SEQUENCES:  35
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;     STREET:  2100 Pennsylvania Avenue, N.W.
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20037
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.[illegible]
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/408,095
;     FILING DATE:  21-MAR-1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Mack, Susan J.
;     REGISTRATION NUMBER:  30,951
;     REFERENCE/DOCKET NUMBER:  P-[illegible]
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202)293-7060
;     TELEFAX:  (202)293-7860
;   INFORMATION FOR SEQ ID NO:  22:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1386 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  214..1155
US-08-408-095-22


alignment_scores:
            Quality:  962.50              Length:    322
              Ratio:  3.540                 Gaps:      0
 Percent Similarity:  99.751   Percent Identity: 98.447%

alignment_block:
"US-09-174-937-5 x "US-08-408-095-22"  ..

 Align seq 1 to:  US-08-408-095-22  from: 1  to: 1386

       20 SLLArgVal3lyMetArgPlaPr PL.PI PLIGlyValLeuSerProSe 64
          ||||||      ||| ||||||  ||||||||||||||||  |||
      177 SAASAPLIIIPPAIAPIIIPAI*IPPAPII*AI PIPLLPAAIPA AIPP 269

       65 rALgAAlgAspMetArgProGlu...rroPrcAsnSerLeuAspLeAsnA 99
          ||  ||||||  |||||||||     |||| |||||||||||||||||||
      270 ACIAPPIAIAACIIPAIPAPAPI*.IPPI.*.P.PPI*PPI*.AIIIAAIG 349

       99 apIhrHisProArgArgIlcLysLc.IhrHiaIlcA.cIlcAsnLcuSer 71
          |  |||||||||||||||||||||||  |||||||||||||||||||||||
      350 PAAATAAATTPAPAAAT  AAA *PAIPPPI.PAPACPIIAPI*.PAIA 716

       72 LeuAspSlaSerSlcSlyLeuIlc.rrIlcAspAspAsnLeuAspSerPr 46
          ||||||  ||||||||||||||||  |||||||||||||||||||||||||
```

```
    397 CTGGATCCTAGTGATGGCTCTGTATTGTCAGATGAT...TTGGATGAAAG 443

     88 oAspGluIleAspIleAsnValAspGluLeuAspThrProAspGlu...A 104

    444 TGGGGAGATTGACTTA.....GATGGCTTAGACACACCGTCAGAGAATA 487

    104 laAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspSer 120

    488 GTAATGAGTTTGAGTGGGAAGATGATCTTCCAAAACCCAAGACTACTGAA 537

    121 GlyGlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgGl 137

    538 GTAATTAGGAAAGGCTCAATTACTGAATACACAGCAGCAGAGGAAAAAGA 587

    137 uAspAsnArgLeuTrpMetThrValValIleGlyGluGlnGluGlnArgI 154

    588 AGATGGACGACGCTGGCGTATGTTCAGGATTGGAGAACAGGACCACAGGG 637

    154 leAspMetLysValIleGluProTyrArgArgValIleSerHisGlyGly 170

    638 TTGATATGAAGGCAATTGAACCCTATAAAAAAGTIATCAGCCATGGG... 684

    171 AspSerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValPheAlaAl 187

    685 .     GGATATTATGGGGATGGATTAAATGCCATTGTTGTATTTGCTGT 728

    187 aCysPheLeuProAspSerSerArgAlaAspTyrHisTyrValMetGluA 204

    729 CTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAGATACCTGATGGACA 778

    204 snLeuPheLeuTyrValIleSerThrLeuGluLeuMetValAlaGluAsp 220

    779 ATCTTTTTAAATATGTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAAC 828

    221 TyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetProGl 237

    829 TACATGATAGTTTATTTAAATGGTGCAACAACTCGAAGAAAAATGCCCAG 878

    237 yLeuGlyTrpMetLysLysCysTyrGlnMetIleAspArgArgLeuArgL 254

    879 TCTGGGATGGCTCAGGAAATGTTATCAGCAAATCGATAGAAGGTTACGGA 928

    254 ysAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThr 270

    929 AAAATCTAAAATCCCTAATCATTGTACATCCTCTTGGTTTATCAGAACA 978

    271 IleLeuAlaValThrArgProPheIleSerSerLysPheSerSerLysIl 287

    979 CTTCTGGCTGTTACAAGACCATTTATTAGCTCGAAATTCAGCCAAAAAAT 1028

    287 sLysTyrValAsnSerLeuSerGluLeuSerGlyLeuIleProMetAspC 304

   1029 TAGATACTTGTTTAATTTGGCAGAACTAGCAGAACTTGTCCCCATGGAAT 1078

    304 ysIleHisIleProGluSerIleIleAsnIleAspLeuLysLeuLysGlu 320

   1079 ACTTTGGCATACCAGAATGCATAAAACAAGTTGATCAAGAACTTAATGGA 1128

    321 Lys 321

   1129 AAA 1131

seq_name ... typical protein data from US_COMB Seq US-08-741-134-5

seq_documentation_block:
; Sequence 5, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,193
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJE-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1236 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1236
US-08-741-134-5


alignment_scores:
            Quality.   90.50              Length:    144
              Ratio:  1.257                Gaps:      5
 Percent Similarity:  50.000   Percent Identity: 26.389

alignment_block:
 US-09-174-937-5 x US-08-741-134-5   ..

 Align seg 1/1  to: US-08-741-134-5  from: 1  to: 1236

      23 GluAspValGlyMetAspIleProPheGluGluGlyVal.....Le 66

     325 GAAGACGATGACGAGGACGATGAGTTCCAAGAATCCGTCCTTTTGACTTT 374

      36 uSerProSerAlaAlaAspMetArgProGluProProAsnSerLeuAspL 53

     375 ATCTCCGGAAGCC.............  CAATACCAACAATCTTTGGACT 409

      53 euAsnAspThrHisProArgArgIleLysLeuThrAlaProAsnIleAsn 69

     410 TGACCATTACTCCAGAAGAAGAATTTAATTCATTGTCACTGGT..... 453

      70 LeuSerLeuAspGlnSerGluGlySerIleLeuSerAspAspAsnLeuAs 86

     454 .       .......TCTTACGCTATCTCCTTGAGCGGTAACTATGTTAA 488

      86 pSerProAspGluIleAspIleAsnValAspGluLeuAspThrProAspG 103

     489 GCATCCATTTGATACTCCAATGGGAGTCCAAGCTGAAGACGAAGATGAAG 538

     103 luAlaAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLys... 118
```

```
      539 ACGGTGACATCTATGACAGTGAAGACTACGACTTGACCCCAGATGAGGAT 588

      119                          ...  ...AspSerGlyGlnGluSerGl 125

      589 GAAATTATTGGCGACGACATGGACGACTTGGATGACGAAGAGGAAGAAGA 638

      125 uSerIleProGluTyrThrAlaGluGluGluArgGluAspAsnArgLeuT 142

      639 AGTTCGTATTGAAGAAGTCCAAGAAGAAGATGAAGAAGATAATGAT.... 684

      142 rpMetThrValValIleGlyGluGlnGluGln 152

      685 .....            GGAGAAGAAGAACAA 699

seq_name: /cpp2_6/ptidata/1/ina/5D_COMB.seq:US-08-500-857A-7

seq_documentation_block:
; Sequence 7, Application US/08500857A
; Patent No. 5912156
;  GENERAL INFORMATION:
;    APPLICANT:  OHATA, SHOZO
;    APPLICANT:  USAMI, STORU
;    APPLICANT:  BURNELL, JAMES N
;    TITLE OF INVENTION:  POLYPEPTIDE HAVING COLD-STABLE PYRUVATE,
;    TITLE OF INVENTION:  ORTHOPHOSPHATE DIKINASE ACTIVITY, DNA ENCODING THE SAME
;    TITLE OF INVENTION:  AND RECOMBINANT VECTOR AND TRANSFORMED PLANTS CONTAINING
;    TITLE OF INVENTION:  THE DNA
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH AND BIRCH, LLP
;      STREET:  8110 GATE HOUSE ROAD SUITE 500 EAST
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/500,857A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  760-209P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2722 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  48..2702
US-08-500-857A-7


alignment_scores:
            Quality:   88.00              Length:    374
              Ratio:  0.515                Gaps:     19
 Percent Similarity:  45.722   Percent Identity: 21.390

alignment_block:
 US-09-174-937-5 x US-08-500-857A-7   ..

Align seq 1/1  to: US-08-500-857A-7  1:GAP: 1  to: 2722

        3 GluGluThrGluPheLeuGluGlyGlyThrArgIleSerArgProAsnSe 19
          ||:::  ::::::  :  :  :  :  ::::::    ::
     1032 GAAAGAAAGAAACTTTACATGCTCCAAACTAGAAATGGAAAGATGAATGT 1081

       19 yLeuLeuSerGluAspValGlyMetAspIleProPheGluGluGlyVall 38
          :    :::    ::: | ::: :  :::    :    :  ::::
     1082 AACTGCTACTGTCAGAACACTGACTCTATG...GTTGAAGAAGCTTTA 1128

       36 euSerProSerAlaAlaAspMetArgProGluProProAsnSerIleAsp 52
          :::::  ...  ::  ::  :  :  :::  :::
     1129 TTACAAAAGAACAAGCCATTATGAGA...ATTGCACTACAATTAGTTGAT 1175

       53 ......LeuAsnAspThrHisProArgArgIleLysLeuThrAlaProAs 67
             ::::::  :::  :  :::
     1176 CAATTACTTCATAAGAATATGCCAGCTAATTAT...GCAGAAGCTGCATT 1222

       67 nLeuAsnLeuSerLeuAspLeuLeuLysGluLeuArgAspA 84
          ::::::  :::  :  :  :::  :::
     1223 AGTTAAA...GGACTTCCACCATCTCCTGAGGTGCTACAGGAGCTTTT 1269

       84 snLeuAspSerProAspGluIleAsp.....................IIe 99
          :::  :::  :::  ::::::                        :::
     1270 TTTTTCATGCCGATGATGCAGTTAAAGAACTTAACTATAGAAAATTCTT 1319

       94 AsnValAspGluLeuAspThrProAspGluAlaAspSerPheLeuTyrTh 111
          :::  :  :  :  ::::::  :::
     1320 CTTCTTAGAGAAGAAACTAAACCTGAAGATTTTGATTCATTCTTTTTT... 1367

      110 rGlyHisAspProThrAlaAsnLysAspSerGlyGlnGluSerIleSer 127
                                                :::  :::
     1368 ..................................GTTGAAGATA 1397

      127 leProGluTyrThr.AlaGluGluGluArgGluAspAsnArgLe..... 141
          ::          :::  :  :  :  :::  :::
     1378 TT......TTAACCTGCAGAGGAGGAAAAAATACACACGCAGTCTCTCTT 1421

      142 ..........TrpMetThrValValIleGlyGluGlnGluGln...Arg 154
                    | :::::  :  :::  :  :::  :::
     1422 GCTAGAGGTATGGGTAAACCATCTCTTCAGCTGTCAACAATTTAAATT 1471

      154 eAspMet...LysValIleGluProTyrArgArgValIleProHisGly 17
          :::::  ::  :  :  :  :::  :::
     1472 TGATGTTGCTAAGAAAATTGCTAAGATTGGAAGTCTTGAAGTTCAAGAA 1521

      170 lyAsp..... 173
          :: :
     1522 GAGATATTTAACTATTGATGCATCAATTGGATGTGTCTATAAGTAGAA 1571

      172 .........................SerGlyTyrTyrGly..... 186
                                  :  :  :::
     1572 GTTGCATTACAACAACCACAACTTGCATCACATTTGCCAACCATTT 1621

      187 .........AspGlyLeuAsnAlaLeuLeuValPheAlaValPheLeu 204
                   :::  :::::  :  :  :  :  :  :  :  :
     1622 AAAATGGGCCAATGAAATTAAAAAGATTGAGTTTTGCTTGTTGAAGATT 1671

      190 euProAspSerSerArgAlaAspLysHisTyrValMetLeuAsnLysPhe 207
          ::::::::::::::::  :::::  :::::
     1672 TTCCATCAGCTGCTAACAAAGCCGTTCAATTCGAGTTGAAGGTATTGGA 1721

      207 Leu.............TyrValIleSerThrLeuGluLeuMetValAlaL 224
          ':'               ::::::::::::::::  :::  :::
     1722 CTTTGCAGAACTCAACGTATCTTGAATGCATTGAAAGACTTCGAATTGT 1771

      219 uAspTyrMetIleVal.............................TyrLeuA 237
                 ::'|':::
     1772 TGTGAAGATGATTCTTTTGAATTATTTGAAGAACTATCTTTA 1821

      227 snGlyAlaThrProArgArgArgMetProGlyLeuGlyTrpMetLysLys 244
          |    :'|   :::::::::  :::  :  :::
     1822 ATGAACTTATGCCACTTCAAAAACAAGATTTCATTGGATTATTGAAGACT 1871
```

```
     244 ..................CysTyrGlnMetIleAspArgArgLeuArgLy 254
                            :::::::::::||          |||::::::
    1872 AIGAAIGGACIIGCAGICACIGTCAGACTTGTTTATCCACCATTACATGA 1921

     254 sAsnLeuLysSerPheIleIleValHisProSerTrpPheIleArgThrI 271
          :                                  ||||::::  ||||:
    1922 A.....................................TTCCTCCCAACTC 1935

     271 leLeuAlaValThrArgProPheIleSerSerLysPheSerSerLys... 286
          ::      :::   ||.   :::::::::    :||:::||||:|||
    1936 TTGAAGAGTTAATGAGAGAAAICTTTGAAATGAAACTTTCAGGTAAGACT 1985

     287 ............................IleLysTyrValAsnSerLeuSe 294
                                    :::||   |||::::::|||
    1986 AAAIAACIIIGAGAAAAAGAAGTTGTTTTAAGTAACTTAAACAACTTAT 2035

     294 rGluLeuSerGlyLeuIle 300
          |*:::::::   :::|||
    2036 GGAAGTTAATCCAATGATT 2054

seq_name: [illegible]

seq_documentation_block:
; Sequence 1, Application US/08/xxx,271
; Patent No. 5920000
;  GENERAL INFORMATION:
;    APPLICANT:  JONES, JONATHAN D
;    APPLICANT:  HAMMOND-KOSACK, KIM E
;    APPLICANT:  THOMAS, COLWYN M
;    APPLICANT:  JONES, DAVID A
;    TITLE OF INVENTION:  PLANT PATHOGEN RESISTANCE GENES AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD, 8TH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/666,271
;      FILING DATE:  19-SEP-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB94/02812
;      FILING DATE:  23-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326428.1
;      FILING DATE:  24-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9409363.0
;      FILING DATE:  11-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  620-7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4091
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3905 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  898..3489
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  967..3486
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  898..966
US-08-666-271-1


alignment_scores:
            Quality:   87.50            Length:    283
              Ratio:  0.788              Gaps:      15
 Percent Similarity:  39.223   Percent Identity: 30.141

alignment_block.
 US-09-174-937-5 x US-08-666-271-1   ..

 Align seg 1/1  to: US-08-666-271-1  from: 1  to: 3905

      68 IleAsnLeuSerLeuAspGlnSerGluGlySerIle...LeuSerAspAs 95
         :::||||||||||   :::      ||||||   .||           ::
    2396 ITGAACITTITCTCAAATGTCTTTGTA[illegible]

      83 pAsnLeuAspSerProAspGluIleAspIleAsnValAspGluLeu..... 112
         :||||||::      :::::::::::|:::::    :::::::::
    3096 AAATTTATCAGTACTCGAATCTTTCAATCTTTCTTCTAATAAAATCAG[illegible]

      99 ..AspThrProAspGluAla..................................  1 4
           :::   |||.....:
    3086 GAGAAATTCCGCAGCAGCTTGCATCCCTCACATTCCTTGAAGTCTTAAAT 3135

     109   .... ..... .          .............................As 136
                                                            :
    3136 CTCTCTCACAATCATCTTGTTGGATGCATCCCCAAAGGAAAACAATTTGA 3185

     105 pSerPhe         .GluTyrThrGlyHisAsp.................P 134
         |||||||         :::|||   ||::::||                  :
    3186 TTCGTTCGGGAACACTTCGTACCAAGGGAATGATGGGTTA[illegible]

     114 roThrAlaAsnLysAspSerGlyGlnCluSerGluSerIlePro...Gln 143
         ||  ......      .:|||.........  .:::  |  .  .   :
    3236 CATTCTCAAAACTTTGTGGTGGTAAACCCAACTGACAACCCAGTTAA 3285

     130 TyrThrAlaGluGluGluArgGluAspAsnArgLeu.....TrpMetLe 144
                 ||||||||  ||||||...  ...        ::::
    3286 CTAGATCAAGAAGAGGAGGAAGAAGATCACCAATGATCAGTTGGCAGGA 3335

     144 rValValIleGlyGluGlnGluGlnArgIleAspMetLysValIleGlu 161
         |||.....:|||
    3336 GGTTCTCGTGGGT.................................... 3348

     161 roTyrArgArgValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsp 177
                                                 |||||
    3349 ..................  .  ...        .  .....TACGGTTGT 3387

     178 GlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSer.. 194
         |||||       .:|||   ::  ....  ..:::      ||:
    3358 GGACTT......GTTATTGGACTGTCCGTAATATACATAAIGTGGTCAAC 3401

     194 ..............     .SerArgAlaAspTyr.................. 208
                           ||||||   |||
    3402 TCAATATCCAGCATGGTTTTCGAGGATCGATTTAAAGTTGTAACATATAA 3451

     198 .........    . .  ................................. 220

    3452 TTACTACGAAAATGAAAAAGCACAAGAAAAGATATTAGTGAGTAGTTATA 3501

     199 .  . .......     .HisTyrValMetGluAsnLeuPheLeuly 228
                         |||'|':::   ||:::  .|  '  :
    3502 CCTCCAGGTATTCCACTTGATCATTATCTTTCAGAAGATTATTTTTTGTA 3551
```

```
     208 rVal  ...          ... . ... .....IleSerThrLeuGluL 215
         |:::                              |||   |||:::|
    3552 IAICGAIGAAAIIAICGACCICCIICAICCICAAAGCICIIAACTTICAC 3601

     216 eaMetValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrPro 231
         |:::::::  |||::::::   |||
    3602 TCTTCATTTTTGAAAATTTCAGGATT..          ..... . . . 3627

     232 ArgArgArgMetProGlyLeuGlyTrpMetLysLysCysTyrGlnMetIl 248
         ::::|||:::|||:::                          |||::::::
    3628 ...CAAACATTTCCGAGT      . ...        ...TCCCAGTTGCT 3653

     248 eAspArgArgLeuArgLysAsnLeuLysSerPheIleIleValHisPro 264
         :    ::::::::      |||   |||:|:|||:::   :::|||
    3654 TGGGAIGCAGAIAAAAGCCTTTTTA...ICTTTCATAGTTTCTTATCCT 3699

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-741-940-1

seq_documentation_block:
; Sequence 1, Application US/07741940
; Patent No. 5952779
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/741,940
;      FILING DATE:  19920109
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.036074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9111
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9605 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  DP2.5(APC)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  34..8562
```

```
US-07-741-940-1


alignment_scores:
            Quality:   87.00              Length:     17
              Ratio:   0.967               Gaps:      1
 Percent Similarity:   51.136   Percent Identity: 22.7..

alignment_block:
 US-09-174-937-5 x US-07-741-940-1    ..

 Align seq 1/1  to: US-07-741-94..-1  from: 1  to: 960.

      49 AsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLe.ThrAl 68
         |'|| ::::::  '|'|  |  :::  :::  ::::::::::::::
    2926 AATAGTGTCACTAGTAATGATGCTTATCCTAAAACAICTCAAATG...AA 2972

      69 aProAsnIleAsnLeuSerLeuAspGlnSerGlyGlySerIleLy.Ser. 91
         | |:::|'|:::        :::::::::  ::::::::::
    2973 ACCCTGATTGAATCCTATTCTGAAGATGATGAAAGTAAGTTTTT'AATT 3022

      92 ..          ............AspAspAsnLeuAspSerProAspL.I. 1
                                  ::::::      ::::::
    3023 AIGGTCAAATACCCAGCCGACCTA'''ATAAAATACATAGT'GAAAT'AT 3072

      91 IleAspIleAsnValAspGluLeuAspThrProAspIleAlaAspSerPh 117
         ::::'   ||    | '||          :::      ::
    3073 ATGGATGATAATGATGCACAACTAGATACACAGCAATAAATTAT...AGTCT 3122

     107 eGluTyrThrGly  .... ..............HisAspProThrAlaA 117
         ::::|'::::                      ::::::   :::
    3120 IAAAIAIICAGATGAGCAGTTGAACTCTGGAAGCCAAAGTCCTT'ACAGA 3219

     117 sn..................................................... 117
         ''
    3170 AIGAAAGATGGGCAAGACCCAAA'ACATAAATAIIAGATCAAAGAAAAAGAA 3219

     118 .........LysAspSerGlyGl.Pl.SerPl.SerIleProPl.TyrTh 137
                 ::::::|''   ::::::  :   :::
    3220 AGTGAGCAAAGACAATCAACGAATCAAACTACAACTTATCCTCTCAIAC 3269

     131 rAlaGluGluGluArgGluAspAsnArgIleTyrMetThrValVa.Ile 146
         '     ||::::   :::|'|::::::      :::          :::
    3270 T......GAGAGCACTGATGATAA'ACACCTTA'CTTCCAACAA'TTTT 3313

     148 lyGluGlnGluGlnArgIleAspMetLysValIleGluProTyrArgAsp 164
         | :::'|'|||:                      :::::::
    3314 GACAGCACGAA.................TTCTTTTCCCATA'A'GTCA 3360

     165 ValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsp.......GlyLe 177
         :::::::'|:::::::::::::                 :::      :::
    3346 CGGGGAGCCAATGGTTCAGAAACAAATCGAGTCGTCAAATAT'GAAAT 3395

     179 uAsnAlaIleIleValPheAlaAlaCys 188
         : |       :::      :::
    3396 IAATCAAAAIGIAAGCCAGTCTTTCTCT 3123

seq_name:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-289-949A-1

seq_documentation_block:
; Sequence 1, Application US/08289949A
; Patent No. 5649212
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
```

```
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, LTD
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4598
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/289,548A
;      FILING DATE:  12-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.035341
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  DP2.5(APC)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  34..8562
US-08-289-548A-1


alignment_scores:
            Quality:   97.00              Length:    176
              Ratio:  0.967               Gaps:        8
 Percent Similarity:  51.136   Percent Identity: 22.727

alignment_block:
 US-09-174-937-5 x US-08-289-548A-1

 Align Seq 1/1  to: US-08-289-548A-1  from: 1  to: 9606

       49 AsnSerIle...AspLeuAsnAspThrHisPr.ArgArgIleLysLeuThrAl 65
          ||||||::::::   ||||||  :.. ...:||  .:::..
     2926 AATACTGTCAGTACTAATGATGCTTATGCTAAAAGAGGTCAAATG...AA 2972

       66 aProAspIleAsnLeuSerLeuAspGlnSer3l.GlySerIleLeuSer  81
          :||:::.|||::  :::::::::|.:::::::::  |||
     2973 ACCCTGGATTGMATCCTATTCTGAACATGAAAGTAAGTTTGCAGTT 3022

       82 .....................AspAspAsnLeuAspSerProAspGlu 90
                                ::::::   |||   ::::::
     3023 ATGGTCAAIACCCAGCCGACCTAGCCCATAAAATACATAGTGCAAATCAT 3072

       91 IleAspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPh 107
          :::'|  '||'     |'|'|'|'|'| ||   :::     |||::
     3073 ATGGATGATAATGATGGAGAACTAGATACACCAATAAATTAT...AGTCT 3119

      107 eGluTyrThrGly.....................HisAspProThrAlaA 117
          ::::'| :::                       ::::::||.:::  !
     3120 IAAAIA'TCAGA'GAGCAGTTGAACTCTGGAAGCCAAAGTCCTTCACAGA 3169

      117 sn ..... ..... 117
          ||
     3170 ATGAAAGATGGGCAAGACCCAAACACATAATAGAAGATGAAATAAAACAA 3219

      118 .........LysAspSerGlyGlnGlnSerGluSerIleProGluTyrTh 131
                  .....:|||  ..... !  ...  ' '  !
     3220 AGTGAGCAAAGACAATCAAGGAATCAAAGTACAACTTATCCTGTTTATAC 3269

      131 rAlaGluGluGluArgGluAspAspArgLeuTrpMetThrValValIleG 148
          !      !!!       ' ||:.....:||'  .::       ::::!
     3270 T      GAGAGCACTGATGATAAACAGCTCAATTTCCAACCACATTTT3 3315

      148 lyGluGlnGluGlnArgIleAspMetLysValIleGluProTyrArgAra 164
          ||:::||||||                  ::::::|| ::::
     3316 GACAGCAGGAA                TTTGTTTCTCCATACAGGTCA 3367

      165 ValIleSerHisGlyGlyAspSerGluTyrTyrGlyAsp.....GlyLe 179
                ....:||:.....:::      |||:::     |||::
     3346 CGGGGAGCCAATGGTTCAGAAACAAATCGAGTGGGTTCTAATCATGGAAT 3395

      179 uAsnAlaIleIleValPheAlaAlaCys 188
          :|||   :::   ':: !
     3396 IAATCAAAATGTAAGCCAGTCTTTGTCT 3423

seq_name: /cgn2_6/ptodata/1/ina/5%_%%MB.seq:US-08-452-654-1

seq_documentation_block:
; Sequence 1, Application US/08452654
; Patent No. 5691454
;  GENERAL INFORMATION.
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND, RAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRODEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND SOMATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORECTAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4598
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,654
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA
;      APPLICATION NUMBER:  US 07/741,940
;      FILING DATE:  08-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.035574
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9606 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       CLONE:  DP2.5(APC)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  34..8562
US-08-452-654-1


alignment_scores:
            Quality:   87.00              Length:    176
              Ratio:  0.967                Gaps:      8
 Percent Similarity:  51.136   Percent Identity: 22.727

alignment_block:
 US-09-174-937-5 x US-08-452-654-1   ..

 Align seg 1/1  to: US-08-452-654-1  from: 1  to: 9606

      49 AsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAl 65
         ||||||:::::   ||||||  ...  ..:||   ::::::
    2926 AATAGTGTCAGTAGTAATGATGGTTATGGTAATAGAGGTCAAATG...AA 2972

      65 aProAsnIleAsnLeuSerLeuAspGlnSerGlnGlySerIleLeuSer. 81
         |||:::|||...          :  :  :  :  :  :  :  :  ||| .
    2973 ACCCTCCATTGAATCCTATTCTGAAGATGATCAAAGTAAGTTTTTGCAGTT 3022

      82 ...              ...  ...AspAspAsnLeuAspSerProAspGlu 90
                                  ::::::   |||   :  ::::
    3023 ATGGTCAATACCCAGCCGACCTAGCCCATAAAATACATAGTGCAAATCAT 3072

      91 IleAspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPh 107
         :::|||   .||      ||||||||||||:|||   :::      |||..
    3073 ATGGATGATAATGATGGAGAACTAGATACACCAATAAATTAT...AGTCT 3119

     107 eGluTyrThrGly.          .  .....        HisAspProThrAlaA 117
         ::::.||:::                 ......||:::.   |
    3120 TAAATATTCAGATGAGCAGTTGAACTCTGGAAGGCAAAGTCCTTCACAGA 3169

     117 sn                                .....   ........ .. ..  117
         ||
    3170 ATGAAAGATGGGCAAGACCCAAACACATAAATAGAAGATGAAATAAAACAA 3219

     118 ...      IysAspSerGlyGlnGluSerGluSerIleProGluTyrTh 131
                  :::::.|||   ......|||   ..:   |||   ||||| .
    3220 AGTGAGTAAACACAATCAAGGAATCAAAGTACAACTTATCCTGTTTATAC 3269

     131 rAlaGluGluGluArgGluAspAsnArgLeuTrpMetThrValValIleG 148
         !     ||||::   .:::|'|.....:||   ...          :::|
    3270 T.     GAGAGCACTGATGATAAACACCTCAAGTTCCAACCACATTTTG 3313

     148 lyGluGlnGluGlnArgIleAspMetLysValIleGluProIyrArgArg 164
         ||:::|||||                 ::::::|||||||||
    3314 GACAGCAGGAA               .....TGTRTTTCTCCATACAGGTCA 3345

     165 ValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsp......GlyLe 179
               ......|||.....  .'      :||:..      ||:.
    3346 CGGGGAGCCAATGGTTCAGAAACAAATCGAGTGGTTCTAATCATGGAAT 3395

     179 uAsnAlaIleIleValPheAlaAlaCys 188
         .|||      ...      ..:.  |||
    3396 TAATCAAAATGTAAGCCAGTCTTTGTGT 3423

seq_name: /cgn2_6/ptodata/1/ina/SD_COMB.seq:US-09-370-235A-1

seq_documentation_block:
; Sequence 1, Application US/09370235A
; Patent No. 5910418
;  GENERAL INFORMATION:
;     APPLICANT:  VOGELSTEIN, BERT
;     APPLICANT:  KINZLER, KENNETH W.
;     APPLICANT:  HILL, DAVID E.
;     APPLICANT:  JOHNSON, KAREN A.
;     TITLE OF INVENTION:  ANTIBODIES AND ASSAYS FOR DETERMINING
;     TITLE OF INVENTION:  MUTATIONS IN THE APC GENE
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  BANNER & WITCOFF, LTD.
;       STREET:  1001 G STREET, N.W.
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  US
;       ZIP:  20001
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/370,235A
;       FILING DATE:  01-JAN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KAGAN, SARAH A.
;       REGISTRATION NUMBER:  32,141
;       REFERENCE/DOCKET NUMBER:  [illegible].48698
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 508 9100
;       TELEFAX:  202 508 9299
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9606 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     POSITION IN GENOME:
;       MAP POSITION:  5q21
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  34..8562
US-08-370-235A-1


alignment_scores:
            Quality:   87.00              Length:    176
              Ratio:  0.967                Gaps:      8
 Percent Similarity:  51.136   Percent Identity: 22.727

alignment_block:
 US-09-174-937-5 x US-08-370-235A-1   ..

 Align seg 1/1  to: US-08-370-235A-1  from: 1  to: 9606

      49 AsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAl 65
         ||||||:::::   ||||||  ...  ..:||   ::::::
    2926 AATAGTGTCAGTAGTAATGATGGTTATGGTAATAGAGGTCAAATG...AA 2972

      65 aProAsnIleAsnLeuSerLeuAspGlnSerGlnGlySerIleLeuSer. 81
         |||:::|||...          :  :  :  :  :  :  :  :  ||| .
    2973 ACCCTCCATTGAATCCTATTCTGAAGATGATCAAAGTAAGTTTTTGCAGTT 3022

      82 ...              ...  ...AspAspAsnLeuAspSerProAspGlu 90
                                  ::::::   |||   :  ::::
    3023 ATGGTCAATACCCAGCCGACCTAGCCCATAAAATACATAGTGCAAATCAT 3072

      91 IleAspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPh 107
         :::|||   .||      ||||||||||||:|||   :::      |||..
    3073 ATGGATGATAATGATGGAGAACTAGATACACCAATAAATTAT...AGTCT 3119

     107 eGluTyrThrGly.          .  .....        HisAspProThrAlaA 117
         ::::.||:::                 ......||:::.   |
    3120 TAAATATTCAGATGAGCAGTTGAACTCTGGAAGGCAAAGTCCTTCACAGA 3169

     117 sn                                .....   ........ .. ..  117
         ||
    3170 ATGAAAGATGGGCAAGACCCAAACACATAAATAGAAGATGAAATAAAACAA 3219

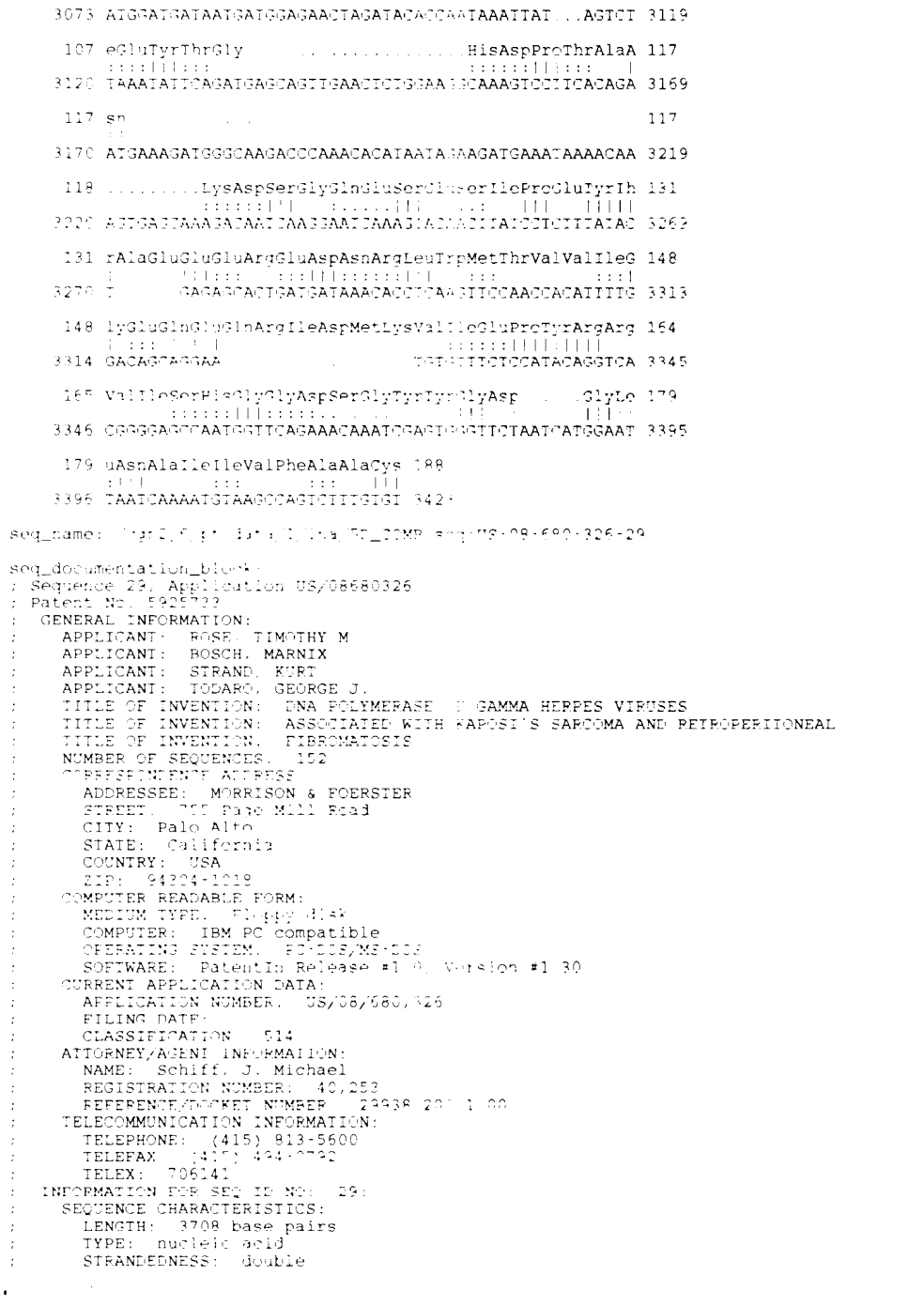
     118 ...      IysAspSerGlyGlnGluSerGluSerIleProGluTyrTh 131
                  :::::.|||   ......|||   ..:   |||   ||||| .
    3220 AGTGAGTAAACACAATCAAGGAATCAAAGTACAACTTATCCTGTTTATAC 3269

     131 rAlaGluGluGluArgGluAspAsnArgLeuTrpMetThrValValIleG 148
         !     ||||::   .:::|'|.....:||   ...          :::|
    3270 T.     GAGAGCACTGATGATAAACACCTCAAGTTCCAACCACATTTTG 3313

     148 lyGluGlnGluGlnArgIleAspMetLysValIleGluProIyrArgArg 164
         ||:::|||||                 ::::::|||||||||
    3314 GACAGCAGGAA               .....TGTRTTTCTCCATACAGGTCA 3345

     165 ValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsp......GlyLe 179
               ......|||.....  .'      :||:..      ||:.
    3346 CGGGGAGCCAATGGTTCAGAAACAAATCGAGTGGTTCTAATCATGGAAT 3395

     179 uAsnAlaIleIleValPheAlaAlaCys 188
         .|||      ...      ..:.  |||
    3396 TAATCAAAATGTAAGCCAGTCTTTGTGT 3423
```



```
; TOPOLOGY: linear
US-08-680-326-29

alignment_scores:
            Quality:   86.00              Length:    241
              Ratio:  0.729                Gaps:     12
 Percent Similarity:  48.963   Percent Identity: 20.747

alignment_block:
 US-09-174-937-5 x US-08-680-326-29

 Align seq 1/1  to: US-08-680-326-29  from: 1  to: 3708

seq_documentation_block:
; Sequence 29, Application US/08680326
; Patent No. 5925720
; GENERAL INFORMATION:
;    APPLICANT: ROSE, TIMOTHY M
;    APPLICANT: BOSCH, MARNIX
;    APPLICANT: STRAND, KORT
;    APPLICANT: TODARO, GEORGE J.
;    TITLE OF INVENTION: DNA POLYMERASE I GAMMA HERPES VIRUSES
;    TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
;    TITLE OF INVENTION: FIBROMATOSIS
;    NUMBER OF SEQUENCES: 152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: MORRISON & FOERSTER
;      STREET: 755 Page Mill Road
;      CITY: Palo Alto
;      STATE: California
;      COUNTRY: USA
;      ZIP: 94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/680,326
;      FILING DATE:
;      CLASSIFICATION: 514
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Schiff, J. Michael
;      REGISTRATION NUMBER: 40,253
;      REFERENCE/DOCKET NUMBER: 23938-20[illegible] 1.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (415) 813-5600
;      TELEFAX: (415) 494-0792
;      TELEX: 706141
;  INFORMATION FOR SEQ ID NO: 29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 3708 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
```

```
seq_name: /cgn2_6/ptodata/1/ina/backfiles1 seq:5171843-8

seq_documentation_block:
;Patent No. 5171843
;     APPLICANT: NUSSENZWEIG, VICTOR
;     TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
;PURIFYING IT
;     NUMBER OF SEQUENCES: 13
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/175,112
;       FILING DATE: 30-MAR-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 754,645
;       FILING DATE: 9-JUL-1985
;       APPLICATION NUMBER: 115,634
;       FILING DATE: 26-OCT-1987
;       APPLICATION NUMBER: 649,903
;       FILING DATE: 12-SEP-1984
;SEQ ID NO:8:
;       LENGTH: 1137
5171843-8


alignment_scores:
            Quality:   84.50              Length:    157
              Ratio:  1.006                Gaps:      9
 Percent Similarity:  53.503   Percent Identity: 27.389

alignment_block:
 US-09-174-937-5 x 5171843-8    ..

 Align seq 1/1  to: 5171843-8  from. 1  to. 1137

      10 LeuGlyThrArgIleSerArgProAsnGlyLeuLeuSerGluAspValGl 26
         ::::||         ||||||   :::|||:::                ||
     157 GTAGGACAAAGTGCTAGCCGAGGCAGAGGCCTT.  ..........  GG 191

      26 yMetAspIleProPheGluGluGlyValLeuSerProSerAlaAlaAspM 43
         |   :::         |||||||||      :::   :::
     192 CGAGAATCCAGATGACGAGGAAGGAGATGCTAAAAAAAAAAAAGGATGGAA 241

      43 etArgProGluProProAsnSerLeuAspLeuAsnAspThrHisProArg 59
         :::   ||||||   |||                          ||||||
     242 AGAAAGCAGAACCAAAAAAT..  .  .  .  .  .  .  .  CCACGT 267

      60 ArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSerGl 76
         :::   ||||||      |||:::              |||:::::::::
     268 GAAAATAAGCTGAAGCAACCAGGA..........     GACAGAGCAGA 302

      76 uGlySerIleLeuSerAsp.AspAsnLeuAspSerProAspGluIleAsp 92
         ||||:.        .:|||  .::   :. :|||:|    .::   |||...
     303 TGGACAGCCAGCAGGAGACAGAGCAGATGGACAGCCAGCAGGTGATAGAG 352

      93             .  .  IleAsnValAspGluLeuAspThrProAspGluAl 104
                          ::::      :|:::::::.      :::
     353 CAGATGGACAACCAGCAGGAGATAGAGCAGGTCGACAACCAGCAGGAGAT 402

     104 aAspSerPheGluTyrThrGlyHisAspProThrAlaAsnLysAspSerG 121
          ::::    ::     :::::::::  |||   :::  ||||||
     403 AGAGCAGATGGAC.    .AGCCAGCAGGAGACAGAGCAGATGGACAGCC 446

     121 lyGlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgGlu 137
          ||||||:::|||:::::   :::       |||   |||:::
     447 AGCAGGAGACAGAGCAGATG  .GACAACCAGTAGGAGACAGAGCAGATG 493

     138 AspAsnArgLeuTrpMetThrValValIleGlyGluGlnGluGlnArgIl 154
         |||||||:::                ||||||          ||||||
     494 GACAACCAG       .......CAGGTGATA      .....GAGCAGCTGGA 522

     154 eAspMetLysValIleGlu 160
          :::::::::|||||||||
     523 CAACCAGCAGGTGATAGAG 541
```

```
seq_name: /cgn2_6/ptodata/1/ina/backfiles1 seq:US-07-864-475A-4

seq_documentation_block:
; Sequence 4, Application US/07864475A
; Patent No. 5494806
;  GENERAL INFORMATION:
;    APPLICANT:  Segre, Gino V.
;    APPLICANT:  Kronenberg, Henry M.
;    APPLICANT:  Abou-Samra, Abdul-Badi
;    APPLICANT:  Juppner, Harald
;    APPLICANT:  Potts, John T. [Jr.]
;    APPLICANT:  Schipani, Ernestina
;    TITLE OF INVENTION:  PARATHYROID HORMONE RECEPTOR
;    TITLE OF INVENTION:  AND DNA ENCODING SAME
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/864,475A
;      FILING DATE:  04-06-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/691,702
;      FILING DATE:  05-04-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  07910/[illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2010
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-07-864-475A-4


alignment_scores:
            Quality:   82.00              Length:    [illegible]
              Ratio:  1.109                Gaps:      [illegible]
 Percent Similarity:  48.066   Percent Identity: 24.8[illegible]

alignment_block:
 US-09-174-937-5 x US-07-864-475A-4/rev    ..

 Align seq 1/1  to reverse of: US-[illegible] 864-475A-4  from: 1  to: 1[illegible]

      83 AspAsnLeuAspSerProAspGluIleAspIleAsnValAspGlu..... 97
         |:::::::::|||:::   |||:| :::::                 :::
    1179 GACGATATTGATGAAGAGGATGAAATTGCAGTAAATGTAGATGGTGATGG 1224

      98 ....LeuAspThrPro..................................... 110
             ||||:|   ||
    1123 CCACCTGGATGATCCACTTTTTCTTTTTTGTGTTGAGTCTCAGTACCCG 1074

     102        ....AspGluAlaAspSerPheGluTyrThrGlyHisAspPro 114
                    :::::|||:::
    1073 GTGTTGGCCAGGGTAGCTCTGACACT.....................GACCCA 1 42
```

```
     115 ThrAlaAsnLysAspSerGlyGln.....GluSerGluSerIlePr 128
         :::::::::::|||'|'|'|||          |::::  ::.  ||
    1041 CACAGCCACGAAGACAGCGGGCAGACCCCAGCCCAAGACTGTGAAGCCCC 992

     128 oGluTyrThrAla......GluGluGluArgGluAspAsnArgLeuTrpM 143
         |:::              ||||||  ...|||'|||...  :::   :
     991 ACAGGIACTICTICTCTCTGAGAAGAAGGCCATGAAGATGAGGCTGTGCAGG 942

     143 et..................ThrValValIleGlyGluGlnGluGlnArg 153
         ::                  '|'|||:::  |  ||||||  :::
     941 TACAGCCCCTCCACCAGAATCCAGTAGTAGTTGTTGGCCAGGAAGTAAAG 892

     154 IleAspMetLysValIleGluProTyrArgArgValIleSerHisGlyGl 170
           :::       :::|||      |||.||  :::  |||||
     891 ...GAAGAAGGTCACAGCCACCCTGCAGCCCGGTTAGCCGGCAGGGCGG 845

     170 yAsp......SerGlyTyrTyrGlyAspGlyLeuAsnAlaIleIleValP 185
         |       |      |'|||||  |  :::  ::::::
     844 TGGCAGGCGGGGGGGGGGCCTGGGCGATGGCGTGAGCTCCTCCTCGGTG 795

     185 heAlaAlaLysPheLeuProAspSerSerArgAlaAspTyrHisTyrVal 201
         :|      ::::::     :::::: ||::::          ::
     794 AGGCGCTCAGCCTCATCAAGCGTGGCGCCAGAGTAGAGCACAGCGTCCTT 745

     202 MetGluAsn 204
         |||:::
     744 GACGAAGAT 736

seq_name. /cgn2_6/ptodata/1/ina/PC_DOMB.seq.US-08-452-655B-1

seq_documentation_block:
; Sequence 1, Application US/08452655B
; Patent No. 5783666
;  GENERAL INFORMATION:
;    APPLICANT:  ALBERTSEN, HANS
;    APPLICANT:  ANAND  PAKESH
;    APPLICANT:  CARLSON, MARY
;    APPLICANT:  GRUBEN, JOANNA
;    APPLICANT:  HEDGE, PHILIP J.
;    APPLICANT:  JOSLYN, GEOFF
;    APPLICANT:  KINZLER, KENNETH
;    APPLICANT:  MARKHAM, ALEXANDER F.
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  THLIVERIS, ANDREW
;    TITLE OF INVENTION:  INHERITED AND S MATIC MUTATIONS OF APC
;    TITLE OF INVENTION:  GENE IN COLORETIAL CANCER IN HUMANS
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001-4598
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,655B
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,048
;      FILING DATE:  12-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/741,940
;      FILING DATE:  08-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.44964
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8532 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  DP2.5(APC)
US-08-452-655B-1


alignment_scores:
            Quality:   82.00              Length:    176
              Ratio:   0.911                Gaps:      8
 Percent Similarity:   51.136   Percent Identity: 22.159

alignment_block:
 US-09-174-937-5 x US-08-452-655B-1   ..

 Align seg 1/1  to: US-08-452-655B-1  from: 1  to: 8532

      49 AsnSerLeuAspLeuAsnAspThrHisProArgArgIleLysLeuThrAl 89
         |'''''.....   :::|||   ::    :::|||   ::::::
    2893 AATAGTGTCAGTAGTAGTGATGGTTATGGTAAAAGAGGTCAAATG...AA 2844

      65 aProAsnIleAsnLeuSerLeuAspGlnSerGluGlySerIleLeuSer. 81
         |||:::|||:::          :::: .:::|||:::::::::   |||'
    2940 ACCCTCGATTGAATCCTATTCTGAAGATGATGAAAGTAAGTTTTGCAGTT 2890

      82                    AspAspAsnLeuAspSerProAspGlu 97
                            ::::::   ||   ::::::
    2889 ATGGTCAATACCCAGCCGACCTATGTCATAAAATACATATGTCAAATCAT 3.4

      91 IleAspIleAsnValAspGluLeuAspThrProAspGluAlaAspSerPh 107
         :::|||   |||      ||||||| ||||||   :::      |:|::
    3040 ATGGATGATAATGATGGAGAACTAGACACACCAATAAATTAT   AGTCT 3086

     107 eGluTyrThrGly..... .....HisAspProThrAlaA 117
         ....|||...                ..:::::'|:::  :
    3087 TAAATATTCAGATGAGCAGTTGAACTTTGTAAGTCAAAGTCTTTTACAGA 3136

     117 sn.....          .  ..............................  117
         ||
    3137 ATGAAAGATGGGCAAGACCCAAACACATAATAGAAGATGAAATAAAACAA 3186

     118 .       .LysAspSerGlyGluGluSerGluSerIleProGluTyrTh 131
                 :::::::|||   :::::::  :::   |  |    |
    3187 AGTGAGCAAAGACAATCAAGGAATCAAAGTACAACTTATCCTGTTTATAC 3236

     131 rAlaGluGluGluArgGluAspAsnAlaLeuTrpMetThrValValIleG 147
         |      |||.      :|||.....|||   :::        ::::
    3237 T......GAGAGCACTGATGATAAACACGTTCAAGTTCCAACTACATTTTG 3286

     148 lyGluGlnGluGlnArgIleAspMetLysValIleGluProTyrArgArg 164
         ||  .||||||                    ....:|| :: |'
    3287 GACAGCAGGAA                    TTCTTTCTCCATATAATTCA 3332

     165 ValIleSerHisGlyGlyAspSerGlyTyrTyrGlyAsp.....GlyLe 177
         :....|||.........:::          |::::       ::
    3313 CGGGGAGCCAATGGTTCAGAAACAAAICGAGTGGGTTCTAATCATGGAAT 3362

     179 uAsnAlaIleIleValPheAlaAlaCys 188
         :|||      :::      :::   '|
    3363 TAATCAAAATGTAAGCCAGTCTTTGTGT 3390

seq_name. /cgn2_6/ptodata/1/ina/PC_DOMB.seq.US-08-468-248A-4
```

```
seq_documentation_block:
; Sequence 4, Application US/08468249A
; Patent No. 5898148
;  GENERAL INFORMATION:
;    APPLICANT:  Segre et al., Gino V
;    TITLE OF INVENTION:  PARATHYROID HORMONE RECEPTOR AND DNA
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS.
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA.
;      APPLICATION NUMBER:  US/08/468,249A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA.
;      APPLICATION NUMBER:  US 07/864,475
;      FILING DATE:  06-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,702
;      FILING DATE:  04-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00786/071003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..1807
US-08-468-249A-4


alignment_scores:
            Quality:   82.00            Length:    153
              Ratio:  1.108              Gaps:      8
 Percent Similarity:  48.366   Percent Identity: 24.837

alignment_block:
 US-09-174-937-5 x US-08-468-249A-4/rev   ..

 Align seg 1/1  to reverse of: US-08-468-249A-4  from: 1  to: 2010

      83 AspAsnLeuAspSerProAspGluIleAspIleAsnValAspGlu..... 97
         |||:::::|||:::   ||||||:.....   |||       ...:
    1173 GACGAIAIIGATGAAGAGGATGAAGTTGAGCAIAATGGAGGCIAGGATGG 1124

      98 .   IeuAspThrPro...  .  .  .  .  .  .  ..... 101
           ||||||   |||
    1123 GCACCTGGATGATCCACTTTTTGTTCCCGGACGTCAAGTCCCAGCACCCG 1074

     102 ..........AspGluAlaAspSerPheGluTyrThrGlyHisAspPro 114
                 .....|||..:                 ||||||
    1073 GTGTTGGCCAGGGTAGCTCTGACACT.        ... ..GACCCA 1042

     115 ThrAlaAsnLysAspSerGlyGln.    ....GluSerGluSerIlePr 128
```

```
         :::::::::: |:|||||||:              :::   :::
    1041 CACAGCCACGAAGACAGCGGGCAGACCCCAGCCGAACACTGTGAAGTCCC 992

     128 oGluTyrThrAla      GluGlnGlnArgGlnAspAsnArgLeuTrpM 141
         ::::              |||:   :::   : :::   :::   :
     991 ACAGGTACTTCTTCTCTGAGAAGAAGGCCATGAAGATGAGGTGTGTGAGG 942

     143 et..................ThrValValIleGlyGlyGlnGlnGlnArg 159
         ::                  :  :::                  :::
     941 TACAGCCCCTCCACCACAATCCACTACTACTTGTTGCCCAGGAAGTAAAG 892

     154 IleAspMetLysValIleGluProTyrArgArgValIleSerHisGly?? 171
         :::   :  :::||  :  :  :  :  :  :  :  :  :  :  :
     891 ...GAAGAAGGTCACAGCCACCCTGCAGTCTGCTACCCGGAGGAGGCGG 845

     170 yAsp.....SerGlyTyrTyrGlyAspProLeuAsnAlaIleIleValP 185
         |       ||  :  :  :  :  :  :  :  :  :  :  :  :
     844 TGGCAGGCGGCGGGGGCGCCTGGGTGATGGTGCAGCTCCTCCTCGGTG 795

     185 heAlaAlaCysPheLeuProAspSerSerArgAlaAspTyrHisTyrVal 201
         |||       :  :  :  :  :  :  :  :  :  :  :  :
     794 ACGCGCTCAGCCTCATCAAGCGTGGCGGCAGAGTAGAGCACAGCGTCGTT 745

     202 MetGluAsn 204
         :|||:|
     744 GACGAAGAT 736

seq_name  /tmp/_5.ptt$ata,1/1ra,1:*Files1.seq:5248670

seq_documentation_block:
;Patent No. 5248670
;     APPLICANT: DRAPER, KENNETH O.;ECKER, DAVID J.;MIHALETTI,
;CHRISTOPHER K.;CROOKE, STANLEY T.
;     TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES F.R
;INHIBITING HERPESVIRUS
;     NUMBER OF SEQUENCES: 15
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/480,297
;       FILING DATE: 26-FEB-1990
;SEQ ID NO:3:
;       LENGTH: 3666
5248670-3


alignment_scores:
            Quality:   82.00            Length:    297
              Ratio:  0.788              Gaps:      15
 Percent Similarity:  43.882   Percent Identity: 24.895

alignment_block:
 US-09-174-937-5 x 5248670-3/rev   ..

 Align seg 1/1  to reverse of: 5248670-3  from: 1  to: 3666

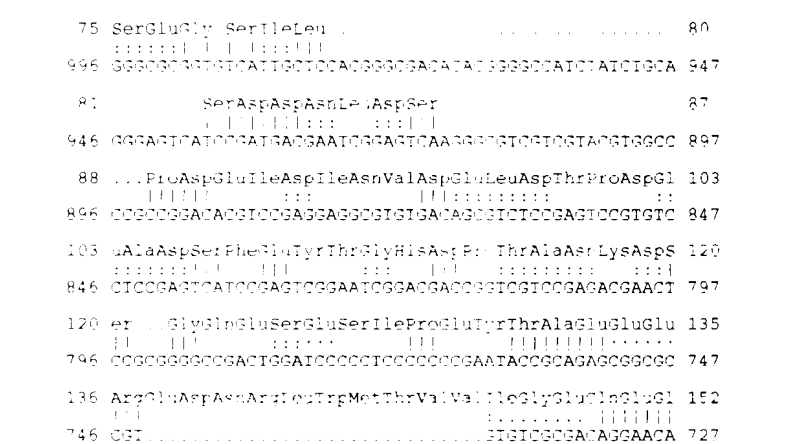
      11 GlyThrArgIleSerArg.........ProAsnGlyLeuLeuSerTrpAs 24
         |||   :||::::::|||           :::
    1219 GGCCTCCGCGTTTCTCGGGCGTGAGTTCCAGGGGGATGG........... 1181

      24 pValGlyMetAspIleProPheGluGlyTyrIleSerProSerAlaA 41
         ||:   :|||:|  :::  :::  :::  :::
    1180 ....GCTAGGCCGTGCCCGTTCCAACGTGGTGGGGTCTAAAGCCC 1140

      41 laAspMetArgProGluProProAspSer..................... 5
         ||||||  :||  |  :::
    1134 CTCCGCGTCGTCCCGGGCCACCGGTGGGATCGTCCGTAAGACCAGTCTG 1 85

      51 .........................LeuAspLeuAspAspThrHisPr 58
                                  :::
    1084 GCCTCTGCGCTCGGCGCCTTGTCCCTTCACCGTTGAGGGAC...TACCG 1 39

      59 eArgArg.IleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGln 74
         ||| |    :::   :::   :::   :::   :::
    1037 GCGTCGGCGCCCGGGCCGGGGGTCTCGGGCAAACATC.........CAG 997
```



```
seq_documentation_block:
; Sequence 1, Application US/07670611
; Patent No. 5330892
; GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Kinzler, Kenneth W.
;    APPLICANT:  White, Raymond
;    APPLICANT:  Nakamura, Yusuke
;    TITLE OF INVENTION:  Gene Mutated in Colorectal Cancer of
;    TITLE OF INVENTION:  Humans
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS
;      ADDRESSEE:  Banner, Birch et al.
;      STREET:  1001 G Street
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20001-4597
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/670,611
;      FILING DATE:  19910313
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.33981
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4181 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  5q21
US-07-670-611-1


alignment_scores:
            Quality:   81.00            Length:    197
              Ratio:  0.771              Gaps:     10
 Percent Similarity:  53.299   Percent Identity: 23.350

alignment_block:
 US-09-174-937-5 x US-07-670-611-1   ..

 Align seq 1/1  to: US-07-670-611-1  from: 1  to: 4181

       1 MetGluGluGluThrGluPheLeuGluGlyThrArgIleSerArgPr 17
         :::::|||||   |||...  ||| ||  ::::::::::|||:::
     530 ATCAGGGAGGAAGATGAGTACTCAGAACTGCGATCAGAACTCAGCCAGAG 579

      17 cAsnGlyLeuLeuSerGluAsp.....                Val3 26
         ...   ......||||||                          ::::
     580 CCAACACGAGGTCAACGAGGACTCTCCGAAGTATGACCAAGACCAGACCT 629

      26 lyMetAspIleProPheGluGluGlyValLeuSerProSerAlaAlaAsp 42
         :::::::|||||  ::::::::::::::          ::::|||||
     630 CTGTCTCTATCCCCGAAAACCAGTCTACCATG...GTTACTGCTGAC 674

      42 MetArgProGluProProAsnSer:L..AspI..AsnAspThrHisPr Ar 59
         |||              |||     |||||||||   ::::::   ::
     674 ATG...........GACAACTGCAGTCAGCCTGAAC...TCAGAACTTCA 7..

      59 gArgIleLysLeuThrAlaProAsnIleAsnLeuSerLeuAspGlnSer3 76
         :|||:::          ||:::      :::     ::::  :
     709 GAGGGTGCTGACAGGGCTGGAGAATGTTGTCTGTGGCAGGAAGAAGAGCA 758

      76 luGlySerIleLeuSerAspAspAsnLeuAspSerProAspGluIleAsp 92
         ..  |||...                 |||      ||||:: |.
     759 GCTGCAGCCTC                TTCCTTGCCGAGGTGGAT 797

      93 IleAsnValAspGluLeuAspThrProAspGlu..................  1 .
                :::|||   ||    ::: .|
     798 AGGCACATTGAGCAGCTCACCACAGCTAGCGAGCACTGTGACCTGGCTAT 847

     104 .      AlaAspSerPheGlu   TyrThrGlyHisAsp......ProT 119
             ....  ....|||        |||.::|||      | |:
     838 TAAGACAGTCGAGGAGATTGAGGGGGTGCTTGGCCGGGACCTGTATCCCA 887

     115 hrAlaAsnLysAspSerGly.............................. 127
                      :::
     888 ACCTGGTTGAAGAGAGGTCTCGGTGGCACTGAGCTGGCTGGGTGAAG 937

     122 GlnGluSerGluSerIleProGluTyrThrAlaGluGluGluArgGluAs 138
         ...|||:::|||||...          ::::::::::|.::::
     938 GAAGAGAATGAGAGCCTGACTGCTATC TTTGTGTCAACTCTGAGAGTCT 987

     138 pAsnArgLeuTrpMetThrVal...ValIleGlyGluGlnGluGlnArgI 154
         ||||||         |||...  :::||    |||:::::::::|||
     988 GAACCGGACTAAGGCCACCATGAATGCCATCCGGGAAGAGCGGGACCGTT 1037
```

```
 154 leAspMetLysValIleGluProTyrArgArgValIleSer 167
     ::      :::|||   |||         |'|:::   |||
1038 TCCGGAGGCGGGTCAGAGAGCTTCAAACTCGACTACAGAGC 1078
```

```
OM of: US-09-174-937-2 *** Issued_Patents_NA.* ***  out_format = pfs

Date: Feb 7, 2000  8:08 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1999 Compugen Ltd.

Command line parameters:
 -MODEL=fram+ -_f2% m!8el -DEV=xlp
 -Q=/cgn2_6/ptodata/1/pubpt/US09174937/rapid_20000207_035108_13295/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCI=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -PGAPOP=6.000 -PGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09174937 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-174-937-2
Query length: 371
Database: Issued_Patents_NA.*
Database sequences: 207303
Database length: 57918364
Search time (sec): 58.270000



score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-409-095-22 -   935.50  2103.84  9.1e-110   1396  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-1 +   114.00   205.36  0.0005     4960  !
/cgn2_6/ptodata/1/ina/5A_1.MB.seq:US-08-[illegible]-3 -   114.00   212.86  0.0002     2195  !
/cgn2_6/ptodata/1/ina/PCTUS9_1.MP.seq:PCT-US95-16983-1 +   114.00   205.36  0.0005     4161  !
/cgn2_6/ptodata/1/ina/PCTUS9_2.MB.seq:PCT-US95-16983-3 -   114.00   212.86  0.0002     2195  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-494-105-15 -   101.50   183.05  0.0091     2004  !
/cgn2_6/ptodata/1/ina/5A_1.MB.seq:US-08-494-105-16 -   101.50   183.05  0.0091     2504  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-1 -    42.50   142.94  1.56      11236  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-2 -    92.00   154.77  0.0375     4119  !
/cgn2_6/ptodata/1/ina/backfiles1.seq:1-- 8-1 -    41.50   144.61  1.24       8224  ! Pate
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-1 -    88.50   141.16  1.09       6316  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-1 -    86.00   148.34  0.7699     3337  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-822B-1 -    88.00   148.34  0.7699     3337  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-[illegible]-170B-1 -    87.00   132.22  6.18       9636  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-459A-2 -    87.00   142.12  1.71       4507  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-826B-3 -    87.00   142.12  1.71       4507  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-[illegible]-119A-1 -    86.50   110.45  99.23     46899  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-[illegible]-141-1 -    86.00   144.89  1.22       3163  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-795-2 -    86.00   135.43  4.09       6316  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-609-4 -    85.50   112.23  78.96     24003  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-[illegible]-255-1 -    85.00   144.00  1.34       2745  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-303-255-13 -    85.00   144.00  1.34       2745  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-646-5 -    84.00   138.02  2.89       3641  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-771-1 -    84.00   142.36  1.66       2616  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-713-637B-6 -    83.50   140.76  2.03       2703  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-96A-1 -    82.00   136.05  2.89       3646  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-781-930-1 -    83.00   138.05  2.89       3646  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-475-647B-5 -    82.00   136.98  3.30       3307  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-257-401A-5 -    83.00   136.98  3.30       3307  !
/cgn2_6/ptodata/1/ina/PCTUS9_2.MB.seq:PCT-US94-09855-1 -    83.00   124.87  15.81      8374  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-210-3 +    82.00   128.39  9.94       5362  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-417-640-9 -    80.50   128.49  8.63       3785  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-6.9-1 -    80.50   100.16  371.54    35935  !
/cgn2_6/ptodata/1/ina/5A_1.MB.seq:US-[illegible]-709-1 +    79.50    97.17  544.92     37895  !
/cgn2_6/ptodata/1/ina/5A_2.MB.seq:US-08-[illegible]-709-10 -    79.50   117.01  42.77      8168  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-08-[illegible]-929-1 -    79.50    97.17  544.92     37895  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-[illegible]-419-10 -    79.50   117.01  42.77      8168  !
/cgn2_6/ptodata/1/ina/5B_2.MB.seq:US-08-[illegible]-765-1 +    79.50   144.75  1.22        289  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-09-098-900-2 +    79.50   150.50  0.5832      633  !
/cgn2_6/ptodata/1/ina/5D_2.MB.seq:US-09-006-877-2 +    79.50   150.50  0.5832      633  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-354-618-1 +    78.50   125.45  14.50      3629  !

/cgn2_6/ptodata/1/ina/5A_1.MB.seq:US-08-[illegible]-1 +    79.50   [illegible]  11.97     3171
/cgn2_6/ptodata/1/ina/5B_1.MB.seq:US-[illegible]-1 -    [illegible]  127.61     128
/cgn2_6/ptodata/1/ina/PCT-US[illegible]-1 -    78.50   127.52  11.11      3095
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-[illegible]-1 -    79.00   128.02  13.35      3164

seq_name: /cgn2_6/ptodata/1/ina/5A_1.MB.seq:US-08-409-095-22

seq_documentation_block:
; Sequence 22, Application US/08408095
; Patent No. 5858678
;  GENERAL INFORMATION:
;    APPLICANT:  Chinnadurai, Govindaswamy
;    TITLE OF INVENTION:  APOPTOSIS-REGULATING PROTEINS
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SUGHRUE, MION, ZINN, MACPEAK & SEAS
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/408,095
;      FILING DATE:  21-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mack, Susan J.
;      REGISTRATION NUMBER:  30,951
;      REFERENCE/DOCKET NUMBER:  [illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)[illegible]060
;      TELEFAX:  (202)[illegible]
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  114..1155
US-08-408-095-22


alignment_scores:
            Quality:  935.50              Length:    948
              Ratio:  3.491                 Gaps:      0
 Percent Similarity:  [illegible]  Percent Identity:  [illegible]

alignment_list:
"DP:09-174-937-2 x US-08-408-095-22"  ..

 Align_seq 1,1  to  "US-08-[illegible]"  from:  1  to:  [illegible]

        5 GlyAlaThrLeuArgMetGlyAsnValAspValLysGlnGlyProAsn 22
          |||:::  |||||||||||||:|:|:|||:|:|:|||:|||:|:|:|
      149 GAATCCAGGGTCAGGATGAAATCTTCAAATCCTAAACAGTAACTCAATA 248

       23 ProLeuProArgProLeuProGlnIleHisGlyValGlyLeuSer 66
          |||||||:|||  |||||||:|:|:|||:|:|:|||:|||:|:|:|
      249 TGAAGATTTTCTCAGCATCAAATTTGAAACATTTAAACAGTAGATA 296

       67 LysSerProValGluAspThrSerSerProProAsnThrLeuAsnProAsn 74
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
      297 TATTAGTTATAAATTGGACCAGACGACAGCAGTTGGCTTACTAGAAGTTAAT 348

       75 GlyAlaHisArgLysArgLysThrLeuValAlaProGlnIleGlnSer 71
          ||  :::::  |||||  ||  :::  :::  :::::::::
```



```
; Sequence 1, Application US/08375300
; Patent No. 5679566
; GENERAL INFORMATION:
;   APPLICANT:  Feng, He
;   APPLICANT:  Jacobson, Allan S.
;   TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;   TITLE OF INVENTION:  THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;   NUMBER OF SEQUENCES:  6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Fish & Richardson
;     STREET:  225 Franklin Street Suite 3100
;     CITY:  Boston
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  02110-2804
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/375,300
;     FILING DATE:  20-JAN-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Fasse, J. P.
;     REGISTRATION NUMBER:  32,983
;     REFERENCE/DOCKET NUMBER:  04020/046001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (617)542-5070
;     TELEFAX:  (617)542-8906
;     TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  4080 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA
US-08-375-300-1


alignment_scores:
            Quality:  114.00              Length:    177
              Ratio:  1.295                 Gaps:      4
 Percent Similarity:  49.718   Percent Identity:  23.729

alignment_block:
 US-09-174-937-2 x US-08-375-300-1   ..

 Align seg 1/1  to: US-08-375-300-1  from: 1  to: 4080

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
    2971 GAGAGTGCTGTTCCTATCGAGTCAATCACCGAAGATGATGAGGATGAAGA 3020

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuG 38
    3021 TGATGAA . . . . . AACGACGATGGTGTCGATTTACTAG 3052

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
    3053 GAGAAGATGAAGACGCGGAGATAAGTACACCGAACACAGAGTCAGCGCCA 3102

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
    3103 GGAAAACATCAGGCAAAGCAA..... 3123

      71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
    3124 ....GACGAAAGTGAAGATGAAGACGATGAGGACGATGACGAGGATGATG 3169

      88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104
```

```
     3170 ACGATGACGATGACGATGATGATGATGATGGAGAAGAAGGCGATGAGGAT 3219

      105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121
          |||:::   :::         :::   :::::   |||   |||:::||
     3220 GATGATGAAGATGATGATGATGATGAGGATGATGATGATGAAGAAGAAGAAGA 3269

      121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
          ':::                          ||||||          :
     3270 CAGC                .          .....GACTCTGATTTGGAGT 3289

      138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154
          ::|||             '||   :::          |||:::::
     3290 ATGGTGGTGATCTTGACGCAGACAGAGATATTCAAATGAAACGAATGTAT 3339

      155 ArgThrVal.........IleIleGlyGluGlnGluHisArgIleAspLe 168
          :::             :::     |||:::|||:::::  :::
     3340 GAAGAGTACGAGAGAAAACTAAAGGATGAGGAAGAAAGGAAAGCGGAAGA 3389

      168 uHisMetIleArgProTyrMetLysValVal 178
           ::::::  |||  ::::::|||||:::
     3390 AGAATTGGAAAGGCAATTTCAGAAAATGATG 3420

seq_name: /tgn1_f/pt1/data/1/fna/5A_COMP_seq/US-08-375-300-3

seq_documentation_block:
; Sequence 3, Application US/08375300
; Patent No. 5679566
;  GENERAL INFORMATION:
;    APPLICANT:  Feng, He
;    APPLICANT:  Jacobson, Allan S.
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
;    TITLE OF INVENTION:  THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. P.
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-375-300-3


alignment_scores:
            Quality:  114.00              Length:    177
              Ratio:  1.295                 Gaps:      4
 Percent Similarity:  49.718   Percent Identity:  23.729

alignment_block:
 US-09-174-937-2 x US-08-375-300-3   ..
```

```
 Align seq 1/1  to: US-08-375-300-3  from: 1  to: 2295

        5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrp[illeg 21
          ||:::::::::::  ::::|||:::::  :::  :::
     1522 GAGAGTGCTGTTCCTATCGAGTCAATCACCGAAGATGATGACGATGAAGA 1571

       21 pGluAspLeuProArgProLeuProGlnGlyThrGlyValGlnLeuLeu 38
          |:::::::              ::::::   '  '''  '''  :::
     1572 TGATGAA....              AACGACGATCCTGTCGATTTACTAG 1603

       38 lySerProValGluAspThrSerSerProProAsnThrLeuAspThrAsn 54
          |:..:     :::       |   '''  '''  :::
     1604 GAGAAGATGAAGACGCGGACATAAGTACACCGAACATAGAGTCAGTGCCA 1653

       55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
          |||   |||..:   ::::::
     1654 GGAAAACATCAGGCAAAGCAA.............................. 1674

       71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
          '''...''''''   :::       ::::  '   '   '   '
     1675 ... GACGAAAGTGAAGATGAAGATGATGATGATGATGACGAGGATGATG 1726

       88 roAspAspLeuAspIleAsnValAspAspIleGluTrpProAspGluThr 104
          '  ''''  |||   :::  '   '   '   '   '   '   '
     1727 ACGATGACGATGACGATGATGATGATGATGGAGAAGAAGGCGATGAGGAT 1776

      105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121
          ||:::   :::         :::   :::::   '   '   :::
     1771 GATGATGAAGATGATGATGATGATGATGATGATGAAGAAGAAGAAGAAGA 1820

      121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
          |:::                          '   '   '   '
     1821 CAGC..............................GACTCTGATTTGGAGT 1840

      138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154
          ::|||             '||   :::          '   '  :::::
     1841 ATGGTGGTGATCTTGACGCAGACAGAGATATTCAAATGAAACGAATGTAT 1890

      155 ArgThrVal.........IleIleGlyGluGlnGluHisArgIleAspLe 168
          :::             :::     '   '   '   '   '  :::
     1891 GAAGAGTACGAGAGAAAACTAAAGGATGAGGAAGAAAGGAAAGCGGAAGA 1941

      168 uHisMetIleArgProTyrMetLysValVal 178
           ::::::  |||  ::::::|||:::::
     1941 AGAATTGGAAAGGCAATTTCAGAAAATGATG 1971

seq_name: /tgn1_f/pt1/data/1/fna/5A_COMP_seq/PCT-US95-16693-1

seq_documentation_block:
; Sequence 1, Application PCT/US95/16693
;  GENERAL INFORMATION:
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED MRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16693
;      FILING DATE:  27-DEC-1995
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4080 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16930-1


alignment_scores:
            Quality:  114.00              Length:    177
              Ratio:  1.295                 Gaps:      4
 Percent Similarity:  49.718   Percent Identity: 23.729

alignment_block:
 US-09-174-937-2 x PCT-US95-16930-1   ..

 Align seq 1/1  to: PCT-US95-16930-1  from: 1  to: 4080

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
         ||:::::::::: ::..|||..... ::: ..:||
    2471 GAGAGTGCTGTTCCTATCGAGTCAATCAGCGAAGATGATGAGGATGAAGA 2520

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuG 38
         |:::::          ::::::  |||||::::|||||||
    3021 TGATGAA..................AACGACGATGGTGTCGATTTACTAG 3052

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
         |:::  :::  ::: ||| ||| |||| ...
    3053 GAGAAGATGAACACGCGGAGATAAGTACACGAACACAGAGTCAGCGCCA 3102

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
         |||  | :::  ::::::
    3103 GGAAAACATCAGGCAAAGCAA.............................. 3123

      71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88
           | |:::|'|||  :::     :::'' ||| |||
    3124 ....GACGAAAGTGAAGATGAAGACGATGAGGATGATGACGAGGATGATG 3169

      88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104
         :|||''  '||  :::  |||||  '|  ||||||
    3170 ACGATGACGATGACGATGATGATGATGATGGAGAAGAAGGCGATGAGGAT 3219

     105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121
         | :::  :::  :::  :::  ||| |||:|||
    3220 GATGATGAAGATGATGATGATGAGGATGATGA.GATGAAGAAGAAGAAGA 3269

     121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138
         |:::                 ||||||        '
    3270 CAGC..................GACTCTGATTTGGAGT 3289

     138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154
         :: '          :|| :::          |||
    3290 ATGGTGGTGATTTGACGAGAGAGATATTGAATGAAACGAATGTAT 3334

     155 ArgThrVal.....IleIleGlyGluGluHisArgIleAspLe 168
         :::          :::  || :::||:::..  ...
    3335 GAAGAGTACGAGAGAAAACTAAAGGATGAGGAAAAAAGGAAAGCGGAAGA 3384

     168 uHisMetIleArgProTyrMetLysValVal 179
          ::::::   |'  ::::::'|:::::
    3385 AGAATTGGAAAGGCAATTTCAGAAAATGATG 3420

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-16930-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9516930
;  GENERAL INFORMATION
;    APPLICANT:  UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL
;    TITLE OF INVENTION:  HETEROLOGOUS POLYPEPTIDE
;    TITLE OF INVENTION:  PRODUCTION IN THE ABSENCE OF
;    TITLE OF INVENTION:  NONSENSE-MEDIATED mRNA DECAY
;    TITLE OF INVENTION:  FUNCTION
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16930
;      FILING DATE:  27-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375,300
;      FILING DATE:  20-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  04020/046WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)542-5070
;      TELEFAX:  (617)542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16930-3


alignment_scores:
            Quality:  114.00              Length:    177
              Ratio:  1.295                 Gaps:      4
 Percent Similarity:  49.718   Percent Identity: 23.729

alignment_block:
 US-09-174-937-2 x PCT-US95-16930-3   ..

 Align seq 1/1  to: PCT-US95-16930-3  from: 1  to: 2295

       5 GluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGlnAs 21
         ||:::::::::: ...|||:::::          :::|:|   ::: |
    1522 GAGAGTGCTGTTCCTATCGAGTCAATCAGCGAAGATGATGAGGATGAAGA 1571

      21 pGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeu 38
         |:::::          ::::::  |||||::::|||||||
    1572 TGATGAA..               AACGACGATGGTGTCGATTTACTAG 1614

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsn 54
         |:::  :::  ::: ||| |||||||||  :::
    1604 GAGAAGATGAAGACGCGGAGATAAGTACACGAACACAGAGTCAGCGCCA 1653

      55 GlyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSe 71
         |||  |||...  .....
```

```
  1654 GGAAAACATCAGGCAAAGCAA............................. 1674

    71 rLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrP 88

  1675 ....GACGAAAGTGAAGATGAAGACGATGAGGATGATGACGAGGATGATG 1720

    88 roAspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr 104

  1721 ACGATGACGATGACGATGATGATGATGATGGATAAGAAGGCGATGAGGAT 1770

   105 AspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAs 121

  1771 GATGATGAAGATGATGATGATGATGAAGATGATGATGAAGAAGAAGAAGA 1820

   121 pThrProValAlaThrAlaLysAsnMetProGlyAspSerAlaAspLeuP 138

  1821 CAGC.....                    ..GACTCTGATTTGGAGT 1840

   138 heGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrp 154

  1841 ATGGTGGTGATCTTGACGCAGACAGAGATATTGAAATGAAACGAATGTAT 1890

   155 ArgThrVal      .     IleIleGlyGluGlnGluHisArgIleAspLe 168

  1891 GAAGAGTACGAGAGAAAACTAAAGGATGAGGAAGAAAGGAAAGCGGAAGA 1940

   168 uHisMetIleArgProTyrMetLysValVal 178

  1941 AGAATTGGAAAGGCAATTTCAGAAAATGATG 1971

seq_name. . 'g'  _ ' '  ]9'.'/1 '.'.'  fA_2[MB.'  ;.US-08-484-105-15

seq_documentation_block:
; Sequence 15, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,115
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59011/RJB/PAC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:

;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2504 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-105-15


alignment_scores:
            Quality:  101.50          Length:    4.
              Ratio:  0.672             Gaps:     1?
 Percent Similarity:  44.282   Percent Identity:  14.941

alignment_block:
 US-09-174-937-2 x US-08-484-105-15   ..

 Align seq 1/1  to: US-08-484-105-15  from: 1  to: 25?4

   59 LysArgLysThrLeuValAlaProGluIleAspIleSerLeuAspGlnSe ??

   92 AGAAGAAAGTCATTGAGGAGTCAACTATTAATTAACGGCATTGATAAA.. 140

   75 rGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuA 12?

  140 .......AGTCTGCTATCTGATGACAGCGCTGACAGTTCTGAT..... 188

   92 spIleAsnValAspAspIleGluThrProAspGluThrAspSerProGlu 138

  176 ....................................ATTGACGAATAGGAA 214

  109 PheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProValAl 157

  191 GTTTACGGTGTTTGGACTGAAGA..TTTTAAAAAAAGAG..GA.GTTT 234

  125 aThrAlaLysAsnMetPro.....................GlyAspS 164

  241 TTATTACAGATCTTTAAAGAAAAATGATGTAATATACGGTTTTGAGAT.. 289

  134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 182

  290 ..................GATATTACTGTACATTATGGAGACTTAATC... 334

  183 GlyArgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAs 231

  320 ......TTTTATCTGGGGGTAATTTGTAAAATTGTACGAAAAAGTAATTGA 369

  197 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrT 244

  364 TAAGCATTCTGGAAAGAAATATGTTAAGCAATTTGTATACGTAGCTT 411

  184 yrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaLysHisLeuPro 2

  414 ATGCTAAGAGAATGGAA.....................ATTAAATT 444

  201 AspSerSerLeuProAspTyrHis.............TyrIleMet..... 241

  440 GAATATTTGTTGCCAGACCGGCATATAAATTATTGTACTTTTTTGTTT 484

  212 .......GluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuAsn 257

  480 TTTTATTAAAAATTTATTTTATTTTTTTTTTTTTTTTTTTTTTTTTA 524

  226 alAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArg 274

  540 CTGAAGCAGAGTTTTTTTCAAAATTTCCCGCTTAATTCGTATAAAACGA 589

  243 ArgMetProGlyIleGlyTrpLeuGlyLysGlyTyrGluGluGlnLeuAspAr 291

  590 AAAGAT..............TTTTTTCGTTGT.................... 619

  259 gArgLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpI 307
```

```
     608 ....................................     AACTTCT 614

     276 heIleArgThrValLeuAlaIleSerArgProPheIleSerValLysPhe 292
         ||||||||      :::   ::::::
     615 TTATCCGACGCGRTGTACACTTGAAA                        640

     293 IleAsnLysIle.....................GlnTyrValHisSerLe 302
         :::|'|||'                      ||:::::||:::||
     641 GTGAACAAATACACAGAAGCCTCTCGATTGGTCTTATTATGCTCATAATCT 690

     302 uGluAspLeuGluGlnLeuIleProMetGluHisValGlnIleProAspC 319
         |''   :::::.                          |||
     691 TGAAAGGATAGAA                    .GATC 707

     319 ysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArgGluSer... 334
           :::::::::   ||||||||:       :::::::::::|||
     708 TTTTGGTTGAGATGGAAGAAAATTTGCGACCAATTAAAAAGAAATCTGGT 757

     335 ............AlaArgProGlnProGluPheValLeuProArgSerGl 347
                     :::|'|||'':::   :.....  '||||||...   ||
     758 TCTAGAGGTCGTGGTCGCCCTCGT...AAATATCCTTTACCAAATGTCGA 804

     347 uGluLysProGluValAlaProValGluAsnArgSerAlaLeu......V 362
         |:::|''   :::   :::   |||:::::::::::       :
     805 AAGCAAAGAAAGCAGTTCCAAAGTTAACTCTAAGGATGAAAATTTTGATT 854

     362 alSerGluAspGlnGluThrSer 369
         ::::::::::|||:::|||:::|||
     855 TACAAGATGATAGTGAATCTTCA 877

seq_name: , ...1_F.p' Data, 1, Lua, BA_COMP seq:US-08-484-106-15

seq_documentation_block.
; Sequence 15, Application US/08484106
; Patent No. 5614618
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1 .., Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-58001/RPO/PAT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2504 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-106-15


alignment_scores:
            Quality:  101.50            Length:    341
              Ratio:  0.672               Gaps:     16
 Percent Similarity:  44.282   Percent Identity: 19.941

alignment_block:
 US-09-174-937-2 x US-08-484-106-15   ..

 Align seg 1/1  to: US-08-484-106-15  from: 1  to: 2504

      59 LysArgLysThrLeuValAlaProGluIleAsnIleSerLeuAspGlnSe 75
         ::|||||||::|||   :::    :::      :::::::''':::
      92 AGAAGAAAGTCATTGAGGAGTCAACTATTAATTAACGGCATTGATAAA.. 140

      75 rGluGlySerLeuLeuSerAspAspPheLeuAspThrProAspAspLeuA 92
              ||||||||||||||||||'      ||:::   :||
     140 .    ..AGTCTGCTATCTGATGACAGCGCTGACAGTTCTGAT....... 175

      92 spIleAsnValAspAspIleGluThrProAspGluThrAspSerLeuGlu 108
                                                 |'|:::   :||'
     176 ....          ...        .              ATTGACGAAGAGGAA 14.

     109 PheLeuGlyAsnGlyAsnGluLeuGluTrpGluAspAspThrProValAl 125
         |||      ...|||      ......   .:::::      ::
     191 GTTTACGGTGTTTGGACTGAAGAGCCC:::AAAAAGAGGCTGGACGTTC 240

     125 aThrAlaLysAsnMetPro.....................GlyAspS 134
         :    ........:                           ||'|||
     241 TTATTACAGATCTTTAAAGAAAAACGATGTAATATATCGCGTTGGAGAT. 289

     134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 167
                    |||   |||:::::::|||||||:::::::::
     290 ................GATATTACTCTACATGATGGAGACTCAAGC... 329

     151 GlyArgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAs 167
         ::::::      ||||||          |||       |||.|
     320    . TTTTATCTGGGGGTAATTTCTAAATTGTACGAAAAAGCAATTGA 363

     167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrT 184
         |   |||      :::   |||:::::::::::::   ::::::     |
     364 TAAGCATTCTGGAAAGAAATATGTTGAAGCAATTTGGTATAGTGGAGCTT 413

     184 yrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPheLeuPro 200
         ||::::::   ::::::                        :::   :||
     414 ATGCTAAGAGAATGGAA      .................ATTAAACCT 456

     201 AspSerSerLeuProAspTyrHis...........TyrIleMet..... 211
         :::      |||||||||   |||          |'::::
     440 GAATATTTGTTGCCAGACGGCATATAAATGAGGTGTAGGTTTGTTGTGT 489

     212 ..     GluAsnLeuPheLeuTyrVal'IleSerSerLeuGluLeuIleV 226
                |||||||||         : :|||:::      ::::::
     490 CCGGGATGAAAAGCTGACTTCATGTAATAGAGCATTGTAATGTGTAAT 539

     229 alAlaGluAspTyrMetIleValTyrLeAsnGlyAlaThrIleArgAsn 242
                 ::::::    ::     |||          :::|''
     540 CTGAAGCAGAGTTTTTTTCAAAATTTCCGGCTGGAATTCCTACAAAACAA 589

     243 ArgMetProGlyIleGlyTrpLeIl I,sCysTyrProMetIleAspAsn 259
         :::                  |||   |||
     590 AAAGAT               TTGTTTCCTTGT..................... 627

     259 gArgLeuArgLysAsnLeuLysSerLeuIleIleValHisProSerTrpS 276
                                                  :::::::|
```

```
      608 ..................  .  ....  ..... AACTTCT 614

      276 heIleArgThrValLeuAlaIleSerArgProPheIleSerValLysPhe 292
          '|||||||       :::   ::::::
      615 TTATCCGACGCGGTGTACACTTGAAA..      .....  .  .....  640

      293 IleAsnLysIle.....  .  ...       ...GlnTyrValHisSerLe 302
          :::|||||||                      |||::::|||:::||
      641 GTGAACAAATACACAGAACCTCTCGATTGGTCTTATTATGCTCATAATCT 690

      302 uGluAspLeuGluGlnLeuIleProMetGluHisValGlnIleProAspC 319
          ||||   ::::|||                                 |||
      691 TGAAAGGATAGAA    ..  .  .....      ............GATC 707

      319 ysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArgGluSer... 334
          :  :::::::::  ||||||:..  .  .  .  :::::::|||
      708 TTTTGGTTGAGATGGAAGAAAATTTGCGACCAACTAAAAAGAAATCTGGT 757

      335 ......      AlaArgProGlnProGluPheValLeuProArgSerGl 347
                   ::::||||||:::   ::::::   |||||||::   ||
      758 TCTAGAGGTCGTGGTCCGCCTCCT...AAATATCCTTTACCAAATGTCGA 804

      347 uGluLysProGluValAlaProValGluAsnArgSerAlaLeu......V 362
          |::::||   ...  ...  |||:::::::::::             :
      805 AAGCAAAGAAAGCAGTTCCAAAGTTAACTCTAAGGATGAAAATTTTGATT 854

      352 alSerGluAspGlnGluThrSer 369
          :::::::.|||...|||..:|||
      855 TACAAGATGATAGTGAAICTTCA 877

seq_name  /tgml_5/ptodata/1/rna/SA_COMB.seq.US-07-853-913-1

seq_documentation_block:
; Sequence 1, Application US/07853913
; Patent No. 5338839
;  GENERAL INFORMATION:
;    APPLICANT:  McKay, Ronald D.G.
;    APPLICANT:  Lendahl, Urban
;    TITLE OF INVENTION:  Nestin Expression As An Indicator of
;    TITLE OF INVENTION:  Neuroepithelial Tumors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,913
;      FILING DATE:  19920319
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,412
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,803
;      FILING DATE:  25-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/201,762
;      FILING DATE:  02-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/180,548
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11236 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-853-913-1


alignment_scores.
            Quality:   92.50            Length:    248
              Ratio:  0.881              Gaps:     19
 Percent Similarity:  44.118   Percent Identity: 24.370

alignment_block:
 US-09-174-937-2 x US-07-853-913-1    ..

 Align seq 1/1  to: US-07-853-913-1  from: 1  to: 11236

       4 ThrGluAlaThrLeuArgMet.........GluAsnValAspValLys.. 39
         |'|:::   |'|'|||||'|          ' ::::::     :::
    8395 ACCCAACAAACACTGAGGGCTGTAGGGAATGAGTAGATGCCAGTGAAGCCC 8444

      17 .GluGluIrpGlnAspGluAspLeuProArgProLeuPro.......... 29
          |||      |||    :::  '  :::  '
    8445 ACCAGAAAAGGTGGATCTAGAGTTAGGAAGCTTCTTGAAAATGAGAAG 8494

      29      ...    ...    ..............................  29

    8495 AAATAGCTAGATCTCTTGGAAAAGAGAATCAAGAGTCACTAGTGTCACTG 8544

      30 GluGluThrGlyValGluLeuLeuGlySerProValGluAspThrSerSe 46
         :::|'|   |||:::|||   :::   ''          :::     ::
    8545 AAAGAAAAAGGTATAGAGACAGTGAAATCTTTAGAAACAGAGATCATAGA 8594

      46 rProProAsnThrLeuAsnPheAsnGlyAlaHisArgLys.......... 59
         :|||   ::::|||   :::  :::      :::'  |
    8595 ACCACTGGAGACTGCAGAAGAGGACCTGGAAAGAAGAGAAGTCTATAGATA 8644

      60 ...  .  .  ...........ArgLysThrLeuValAla 80
                              :::  :::
    8645 CTCAGGAGCCATTGTGGTCTACTGAAGTGGCTAGAGACACTAGAAACTT 8694

      66 ProGluIleAsnIleSerLeuAspGlnSerGluGlySerIleLeuSer.. 81
         '|'|||              |  :::      '   '   '
    8695 CCAGAA .            GATGACCCTCCTCTGAACCCTAGGTTCTGT 8729

      82 .AspAspPheLeuAspThrProAspAspLeu..................... 91
          |||:::      :::|||      :::|'
    8730 GGATGAGAACCGAGAGACACTGACATCCTTGAAAATGAGACTCAAGAAC 8779

      92 ..................  .......AspIleAsnValAspAsp..... 97
                             :::  :::'  :::
    8780 TGAGCTCTCTGGGCAAGTGGAACGTAGAGACCAGGTAGAGGATCATTCAG 8829

      98 ...................      IleGluThrProAspGluThrAspSe 114
                               ::::::     ::::::  :::
    8830 CAGTGCCTGCAAGTAGAAGAGGCTCTCCACCACCAACCACCAAGATC 8879

     106 rLeu...      ...     GluPheLeuGlyAsnGlyAsnGluLeuGlyT 118
         ||||             ||:::   ::::::'  '   :::  :::
    8880 ICTGAGAGAGGTGAAGCAGGAGCTGCTAGCTCTGGAAATCAACAGCGGT 8929

     118 rpGluAsp...          . ..............AspThrProValAla 128
         |||'|:|'                          ::::::  :::
    8930 GGGAGGATGTGGTGGAGGGCAAAGCAGTGGGTCAGGAAGCACTTCTGGCA 8979

     126 ThrAlaLysAsnMetProGlyAspSerAlaAspLeuPhe.....GlyAs 141
         '|':::   .          |'  :::  :::  '          :::
    8980 ACCACAGGAGTGGGAACTGAGGATAAGGCAGAGTTGTATCTGAGGGGTA 9029
```

```
     140 pGlyThrThrGluAspGlySerAlaAlaAsnGl,ArgLeuTrpArgThrV 157
     9060 AGGTGGAGAGGAAGAA...GCTGCAGCAGAGGAGAGCTGTTGCAGGAT. 9075

     157 alIleIleGlyGlu 161
    9076 ..ATTGTGGGGGAG 9087

seq_[illegible] ..rpI_[illegible]gr1:1tn,1,1ne,TA_2IMP s.; US-08-119-125A-3

seq_documentation_block.
; Sequence 3, Application US/08119125A
; Patent No. 5610011
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, Hilda Elizabeth
;    APPLICANT:  VECHT, Uri
;    TITLE OF INVENTION:  DNA Sequences which code for Virulence
;    TITLE OF INVENTION:  Characteristics of Streptococcus suis and parts thereof, polyp
;    TITLE OF INVENTION:  antibodies derived therefrom and the use thereof for the diagn
;    TITLE OF INVENTION:  protection against infection by S. suis in mammals, including
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Centraal Diergeneeskundig Instituut
;      STREET:  Edelhertweg 15
;      CITY:  PH Lelystad
;      STATE:
;      COUNTRY:  The Netherlands
;      ZIP:  NL-8219
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS v.6.0
;      SOFTWARE:  WordPerfect v. 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/119,125A
;      FILING DATE:  20-SEP-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/NL92/00054
;      FILING DATE:  19-MAR-1992
;      APPLICATION NUMBER:  NL 9100510
;      FILING DATE:  21-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Handal, Anthony H.
;      REGISTRATION NUMBER:  25275
;      REFERENCE/DOCKET NUMBER:  SMITHHE11-125
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203) 838-8589
;      TELEFAX:  (203) 838-8794
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4118 base pairs
;      TYPE:  Nucleic acid with corresponding amino acids
;      STRANDEDNESS:  single stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus suis type II (pathogenic)
;    FEATURE:
;      OTHER INFORMATION:  Muramidase released protein (MRP) gene
;    FEATURE:
;      NAME/KEY:  promoter -35 region
;      LOCATION:  bp 4 to 9
;    FEATURE:
;      NAME/KEY:  promoter -10 region
;      LOCATION:  bp 29 to 34
;    FEATURE:
;      NAME/KEY:  promoter -35 region
;      LOCATION:  bp 40 to 45
;    FEATURE:
;      NAME/KEY:  promoter -10 region
;      LOCATION:  bp 63 to 68
;    FEATURE:
;      NAME/KEY:  ribosome binding site
;      LOCATION:  bp 147 to 152
;    FEATURE:
;      NAME/KEY:  signal peptide
;      LOCATION:  bp 159 to 299
;    FEATURE:
;      NAME/KEY:  mature peptide
;      LOCATION:  bp 300 to 3926
;    FEATURE:
;      NAME/KEY:  proline rich region
;      LOCATION:  from bp 2757 to 3014
;    FEATURE:
;      NAME/KEY:  repetitive units
;      LOCATION:  from bp 3015 to 3176, 3423 to 3584 and 3587 to 374.
;    FEATURE:
;      NAME/KEY:  membrane anchor sequence
;      LOCATION:  from bp 3825 to 3926
;    FEATURE:
;      NAME/KEY:  dyad symmetry regions
;      LOCATION:  from bp 4069 to 4090 and from bp 4087 to 4118
US-08-119-125A-3


alignment_scores:
            Quality:   92.00              Length:    305
              Ratio:  0.676                Gaps:     10
 Percent Similarity:  44.590   Percent Identity:  18.689

alignment_block:
 US-09-174-937-2 x US-08-119-125A-3   ..

 Align seg 1/1  to: US-08-119-125A-3  from: 1  to: 4118

      22 GluAspLeuProArgProLeuPr[illegible]GluThrGlyValGluLeuLeuGl 58
    3078 GAAGAGGGAACAAAACCAAACAAATCAATCCCAGGTTACGAGTTCACAGG 3127

      38 ySerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsnG 55
    3128 TAAAACTGTTACTGACGAAGATGGC          ..........AACA 3156

      55 lyAlaHisArgLysArgLysThrLeuValAlaProGluIleAsn..... 80
    3157 CAACTCACATCTACAAGAAAACA.........CCAGAAGTTAAGAATGGT 3197

      70 ..... ...IleSerLeuAspGlnSerGluGlySerLeuLeuSerAspAs 93
    3198 ACAGTTGTTGTTAACTATGTAACAGAAGATGGCACAGTTATCAAGGAACC 3247

      83 pPheLeuAspThrProAspAspLeuAspIleAsnValAspAspIleGluT 109
    3248 TGTAACAGATACACCAACTTCT ........................... 3269

     100 hrProAspGluThrAspSerLeuGluPheLeuGlyAsnGlyAsnGluLeu 119
    3269 .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. .. 3269

     117 GluTrpGluAspAspThrProValAlaThrAlaLysAsnMetPro..... 141
    3270 ... .CCAGAAGGCACACCATA[illegible]ACAACAAACCTAAGAC 3314

     132 ....        . GlyAspSerAlaAspLeuPheGlyAspGlyThrThrG 144
    3314 AATCACTTTCAAAGGTGAAGAGTATCAATTGGTTCGTGTTGACTGTACAG 3363

     144 luAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrValIleIleGly 193
    3364 AAAACGGTAAAGTTGTAGAAGGTGAAACAGTTGTGACTTACGTTTACCGT 3413

     161 GluGlnGluHisArgIleAspLeuHisMetIleArgProTyrMetLysVa 177
    3414 AAAGTCGAA.....  .              ACACCTGCTAAGAAAGT 3448
```

```
seq_name: /cgn2_6/ptodata/1/ina/backfiles1 seq-5180808-1

seq_documentation_block:
;Patent No. 5180808
;     APPLICANT: RUOSLAHTI, ERKKI I
;     TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
;SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
;ANTIBODIES, AND METHODS OF DETECTING THE SAME
;     NUMBER OF SEQUENCES: 4
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/441,179
;       FILING DATE: 27-NOV-1989
;SEQ ID NO:1:
;       LENGTH: 8224
5180808-1


alignment_scores:
            Quality:   91.50              Length:    449
              Ratio:  0.446                Gaps:     21
 Percent Similarity:  45.657   Percent Identity: 19.822

alignment_block:
 US-09-174-937-2 x 5180808-1   ..

 Align seq 1/1  to: 5180808-1  from: 1  to: 8224
```

```
 2316 ...........TTAGACACTAGCAGGATAATCATAGAAAGCTTTTTTGAG 2354

  290 ValLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeuGl 306
                :::          |||      ||   :::
 2355 GTTCCTGCAACCACC.....ATTTATCCAGTTTCTGAACAACCTTCTGC 2398

  306 uGlnLeuIleProMetGluHisValGlnIleProAsp             318
       :::::::::|'   :::   ||:::     ||
 2399 AAAAGTGGTGTTTACTAAGTTTGTAAGTGAAATAGACACTTCTGAGTGGA 2448

  319 ..CysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArg...... 332
                   ||||||:::::    |    ::::::
 2449 TTTCCAGTACCACTGTTGAGGAAAAGAAAAGGAAGGAGGAGGAGGGAACT 2498

  333 ..............................GluSerAlaArgProGlnPr 339
                                    ::: ||.......  ..:
 2499 ACAGGTATGGGTTCTATATTTGAGGTATATTCATCTACACAGAGATCGGA 2548

  339 oGluPheValLeuProArgSerGluGluLysProGluValAlaProValG 356
       ::::::::|'|:||     ..   |||   |     ||||||
 2549 TCAATTAATTTTACCCTTTGAATTAGAAAGTCCAAATGTAGCT.....   2591

  356 luAsnArgSerAlaLeuValSerGluAspGlnGluThrSerMetSer 371
                     :::|||:::|||:::    |||      |||
 2592 .................ACATCTAGTGATTCAGGTACCAGGAAAAGT 2621

seq_name: US-08-808-793-1, Data, DB_COMB Seq:US-08-808-793-1

seq_documentation_block:
; Sequence 1, Application US/08808793
; Patent No. 5858713
;  GENERAL INFORMATION:
;    APPLICANT:  Soderlund, David M
;    APPLICANT:  Ingles, Patricia J
;    TITLE OF INVENTION:  CALCIUM PERMEABLE INSECT SODIUM CHANNELS
;    TITLE OF INVENTION:  AND USE THEREOF
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,793
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/034,361
;      FILING DATE:  24-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/012,649
;      FILING DATE:  01-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Braman, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  19603/1702 (C-1306?)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1636
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-808-793-1


alignment_scores:
           Quality    88.50              Length:     283
             Ratio:   0.651                Gaps:      19
 Percent Similarity   48.057   Percent Identity   22.968

alignment_block:
 US-09-174-937-2 x US-08-808-793-1

 Align seg 1/1  to: US-08-808-793-1  from: 1  to: 6318

   11 GluAsnValAspValLysGluGluTrpGlnAspGlu...AspLeuProAr 26
      |||         .::  ||:::.   .::|||:::        |||    ::
 3402 GAAGAAATCAAAATTCATGAACAACACAACGATGATTGGAAACTCAATAA 3451

   26 gProLeuProGluGluThrGlyVal GluLeu     ........LeuGly 38
      '|||        :::   ||||||:::  ||||||          ||::::
 3452 ACCACCAAGACAATAGACTGGAACATGAGCTAAACCATAGAGGTTTGTCC 3501

   39 SerProValGluAspThrSerSerProProAsnThrLeuAsnPheAsnGl 55
               :::||||||:::|||            :::|||      ||
 3502 ATACAGGACGATGACACTGCCAGC............ATTAACTCATATGG 3539

   55 yAlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSerL 72
       |:::|||:::::|||
 3540 TAGCCATAAGAATCGACCATTC............................ 3561

   72 euAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrPro 88
            ::::::|||:::               |||   :::   ::::::
 3562 ....AAGGACGAGAGCCACAAGGGCAGCGCCGAGACCATCGAGGGCGAG 3606

   89 AspAspLeuAspIleAsnValAspAspIleGluThrProAspGluThrAs 105
      :::      |||::::::   ::::||||:::            ::::|||   ||
 3607 GAGAAACGCGACGTCAGCAAAGAGGACCTCGGCCTCGACGAGGAACTGGA 3656

  105 pSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTrpGlu....... 119
      |:::   |||      |||:::   ......|||:::   :::
 3657 CGAG...GAGGCCGAGGGCGATGAGGGCCAGCTGGATGGTGACATTATCA 3703

  120 ...........................AspAspThrProVal...... 124
                                 ||||||   |||:::
 3704 TTCATGCGCAAAACGACGACGAGATAATCGACGACTATCCGGCCGACTGT 3753

  125 .........AlaThrAlaLysAsnMetPro      GlyAspSerAl 135
                :::       |||:::::|||            ||||||:::
 3754 TTCCCCGACTCGTACTACAAGAAGTTTCCGATCTTGGCCGGCGACGAGGA 3803

  135 aAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGlyA 152
       ...  |||   :.:|||                  :::|||   |
 3804 CTCGCCGTTCTGGCAAGGATGG...............GGCAATTTAC 3835

  152 rgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAspLeu 168
      |||||    :::|||      :::   |||:::::
 3836 GACTG...AAAACTTTTCAATTAATTGAAAATAAA.............. 3867

  169 HisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrTyrGl 185
                                                 |||:::
 3868 ............................................TATTTTGA 3875

  185 yGluGlyLeuAsnAlaIleIleValPheAla       AlaCysPheLeuP 200
        ....      .   |||:::::::::            |||     |||
 3876 AACCGCAGTTATCACTATGATTTAATGAGTAGCTTAGCTTTGGCCTTAG 3925

  200 roAspSerSerLeuPro       ...    AspTyrHisTyrIle 210
        |||      ||||||                  |||     |
 3926 AAGATGTTCATTTACCGATCGACCTGTCATGCAGGATATACTGTACTAC 3975

  211 MetGluAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAl 227
      |||.........:|||       |||::         :|||:|:::    ::::
```

```
    3976 ATGGACAGGATATTTACG.  GTGATATTCTTTTTGGAGATGTTGATC.. 4020

     227 aGluAspTyrMetIleVal.TyrLeuAsnGlyAlaThrProArgArgArg 243
              :::::   :::    !!!:::   ::::::!!!
    4021 ....AAATGGTTGGCCCTGGGCTTTAAGGTTTACTTCACCA..... 4057

     244 MetProGlyIleGlyTrpLeuLys......LysCysTyrGlnMet 256
         ||||||||:::!||||||::::::      :::!||||::::::
    4058 ATGCCTGGTATTGGCTGGATTTCGTGATTGTATGCTATCGCTTA 4102

seq_name: /cgn2_6/ptodata/1/ina/5A_COMP_seq/US-09-072-610-1

seq_documentation_block:
; Sequence 1, Application US/08072610
; Patent No. 5532133
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens,
;    TITLE OF INVENTION:  Monoclonal Antibodies, and Diagnostic Assays
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,610
;      FILING DATE:  19930602
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/07696
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3337 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-072-610-1


alignment_scores:
            Quality:   88.00              Length:    378
              Ratio:  0.506                Gaps:     19
 Percent Similarity:  46.032   Percent Identity: 23.545

alignment_block:
 US-09-174-937-2 x US-08-072-610-1   ..

 Align seq 1/1  to: US-08-072-610-1  from: 1  to: 3337

       8 LeuArgMetGluAsnVal......AspValLysGluGluTrpGlnAspGl 22
         !!|:.....!!!!!!..      :!!!!!  :!!:          ::
     237 TTAGAATTGGAAAATGCTTCTGATGATGTTGTAGAGGTG.........GA 277

  22 uAspLeuProArgProLeuProGluGluThrGlyValGluLeuArgGlyS 71
     !!!                !:::::::   :  :::             :
 278 GGAT..............CCTTCAAACGACGGTTTAGAATTAGAAGAGG 312

  39 erProValGluAspThrSerSerProProAspThrLeuAsnPheAsnGly 88
     ::      :::::::: !!:::      :::  :
 313 AAAATTTTGATGAGAATTCAGGTGATGATGAAACTCTT............. 347

  56 AlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSerLe 71

 351 .       ...          ..............................TT 382

  72 uAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrProA 88
     ||||   :::                :::: :  :  :          :
 383 AGATGCTACCCCC.............GAAGATGATTTTGCCTTAACTA..... 416

  89 spAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr... 104
         !!!:::   :::::: :    : !      :::  :
 417 ........GATTTGCCAATTGAAGATGATGAGGAAGTCAACGAAATGTTA 460

 105 AspSerLeuGluPheLeuGly......AsnGlyAsnGluLeuGluTrpGl 119
     |||:::   :!!   !!!!!!       :::   :::::::::      :
 461 GATGGAGGTGAATCATTAGCAGACGTTTCCACTGAAGATATGGAAACAGA 510

 119 uAspAspThr......ProValAlaThrAlaLysAsnMetProGlyAspS 134
     !!!!   :::        :::      :  :::::::::  :     ::
 511 AGATGGCTCAACAGATGATACGGAAACAGAAGAAGGACTACCTGGTGATA 560

 134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 151
                    :::!!                 !:::
 561 TG.        ...      GAAGGATAAGAAGAAGCTGTCGCATATGGAAGCA 593

 151 GlyArgLeuTrpArgThrValIleIleGlyGlnGlnHisArgIleAs 167
     | |:::          :::          | :::        :::::
 594 GGGGAAGAAGCTGGTGATTTGGAAGCAGGGGAAGAAACTGGCGATTTGGA 643

 167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyr 183
     :                  :::::::::      :::::::
 644 AGCAGGGGAAGAAACTGGCGATTTGGAAGCAGGGGAAGAAACTGGCGATTTGGA 699

 184 TyrGlyGluGlyLeuAsnAlaIlelle..............ValPheAl 195
     :::   :::!!!   :::         ::             :::
 700 TGGAAGCAGGGGAAGAAACTGGCGATTTGGAAGCAGGGGAAGAAACTGGA 749

 195 aAlaCysPheLeuProAspSerSerLeuProAspTyrHisLysIleMetG 212
                |!!      :::::::  :                 :
 750 GATGCGGAAACTGAAGAAGGAGCAACTGGAGATGCGGAAACTGAAAATGG 799

 212 luAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAlaGlu 228
     !!:::!!!   :::   :::   :::   ::::::  :::      :::
 800 AGCAACTGTGTATGTAGACACAGAAGATAGTTCAGCTGATGGAGCAGAAA 849

 229 AspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetPr 245
       !:!!!:::::::
 850 AAGTACATGTTCCTG..................................... 928

 245 eGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgAspLeuA 262
                !!!! !!!!  :  :::::::
 929 .       .... .CTCAAGAAAATGTACAACCTG.................. 849

 262 rgLysAsnLeuLysSerLeuIleIleValHisProSer.............. 274
                          !!:::::::      :
 850      .           . .CTGATAGTAATGATGTTCTTTGGAAGTATTT 882

 275 TrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerValLy 291
     !!:::              !!!    :    :::      :  :
 883 TGGATAAAGATATAATTTTTGATCATATTAAAG.......ATTTCGAGCCA 924

 291 sPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeuGlu...G1 317
     :::  :!!!:::...      :  :::      :::         :::
 927 CTATTCGAACAAATTGTGGCGGTA...CTGCTAAACATGTTACGTGACA 974
```

```
     307 nLeuLeuProMetGluHisValGlnIleProAspLysValLeuGlnTyrG 324
         ||  ||| ||| |||    ||  |||| |          |||       |
     974 AGAATTGCCAATGAAACCTGTACCATTACCA         GTGGCAGAAG 1014

     324 luGluGluArgLeuLysAlaArgArgGluSerAlaArgProGlnProGlu 340
         ||      ||||||   |||||||||   ||||||   |||||   |||
    1015 AGCCCGCGCAAGIACCAGCGGAAGAATIAGAIGCCACTCCAGAGGATGAC 1064

     341 PheValLeuProArgSerGluGluLysProLeuValAlaProValGluAs 357
         | ||||| |     |||||||||      ||  ||      |||   ||
    1065 ITCGCATTAGATGTTACAGAATCICCCGAGGAAGTAGAATTAGTATTAGA 1114

     357 nArgSerAlaLeuValSerGluAspGlnGlu 387
         |||||||||       |||||||||   | |
    1115 TGAAGAGGCAACTGAAGAAGAATCAACGGAA 1145
```

seq_name: /cgn2_6/ptodata/1/ina/60_COMP_seq/US-09-719-822B-1

```
seq_8...Publish_kl.A
; Sequence 1, Application US/09719822B
; Patent No. 5874527
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,822
;      FILING DATE:  09/30/96
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5444/2pct/US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS.
;      LENGTH:  3337 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE.
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-719-822B-1


alignment_scores:
            Quality:   88.00              Length.    378
              Ratio:  0.506               Gaps:      19
 Percent Similarity:  46.032   Percent Identity: 23.545

alignment_block.
 US-09-174-937-2 x US-08-719-822B-1   ..
```

```
Align seq 1/1  to: US-08-719-822B-1  from: 1  to: 3337

      8 LeuArgMetGluAsnVal..    AspValLysGluGluTrpGlnAspGl 22
        |||:::::|||||:::       ||| |   |||              |
    237 TTAGAATTGGAAAATGCTTCTGATGATGTTGTACAGGTC.....GA 277

     22 uAspLeuProArgProLeuProGluGluThrGlyValGluLeuLeuGlyS 39
        ||||               |||:::::   |||::|||||       :
    278 GGAI..............CCTTCAAACGACGGTTTAGAATTAGAAGAGG 318

     39 erProValGluAspThrSerSerProProAsnThrLeuAsnPheAsnGly 58
        ..      ..........:|||:::      :::||||||
    313 AAAATTTTGATGAGAATTCAGGTGATGATGAAACTCTT..........  350

     56 AlaHisArgLysArgLysThrLeuValAlaProGluIleAsnIleSerLe 72
                                                        :
    351   ... ..... ....      ..........................II  382

     72 uAspGlnSerGluGlySerLeuLeuSerAspAspPheLeuAspThrProA 89
        ||||   :::                 :::||||||||||      |||
    353 AGATGCTACCCCC............GAAGATGACTTTGCCTTAACA.... 386

     89 spAspLeuAspIleAsnValAspAspIleGluThrProAspGluThr... 106
              |||:::   ..... :|||:   |||      :::| |||:
    387 ........GATTTGCCAATTGAACACGATGAGGAAGTCAACGAAACGTTA 429

    105 AspSerLeuGluPheLeuGly......AsnGlyAsnGluLeuGluTrpGl 124
        |||:::   |||   ||||||      :::   ::::::::::||   |
    429 GATGGAGGTGAATCATTAGGAGAGGTTTCCACTCAAGATATGGAAACAGA 478

    119 uAspAspThr    . ProValAlaThrAlaLysAsnMetProGlyAspS 134
        ||||  ...      :::  |||  ..:::::::::|||| ||
    479 AGATGGCTCAACAGATGATACGGAAACAGAAGAAGGACTACCTGGTGATA 528

    134 erAlaAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsn 153
                        ::|||     |||   ||:::
    529 TG     ..........GAAGGAGAACAAGAAGCTGGCGATATGGAAGCA 568

    151 GlyArgLeuTrpArgThrValIleThrGlyGluGlnGlnHisArgIleAs 167
        |||...           ..:     ||  ||:::       :::::
    564 GGGGAAGAAGCTGGTGATTTGGAAGCAGGGGAAGAAACTGGCGATTTGGA 613

    167 pLeuHisMetIleArgProTyrMetLysValValThrHisGlyGly.Tyr 185
        :                 :::::::::      ::::::||
    614 AGCAGGGGAAGAAACTGGCGATTTGGAAGCAGGGGAAGAAACTGGTGATT 663

    184 TyrGlyGluGlyLeuAsnAlaIleIle..............ValPheAl 195
        :::  ::|||  :::    ||                   :::
    664 TGGAAGCAGGGGAAGAAACTGGCGATTTGGAAGCAGGGGAAGAAACTGGA 713

    195 aAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrIleMetC 212
              |||      ::::::| |                  ||||
    714 GATGCGGAAACTGAAGAAGGAGCAACTGGAGATGCGGAAACTGAAAATGG 763

    212 luAsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAlaGlu 228
        ||:::|||   :::  :::  :::   ::::::|||:::      :::
    764 AGCAACTGTGTATGTAGACACAGAAGATAGTTCAGCTGATGGAGCAGAAA 813

    229 AspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetPr 245
          |||||:::::
    814 AAGTACATGTTCCTG.....  ...      ..............  828

    245 oGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgArgIleA 262
              |||||||||    :  ::::::
    829  .........  CTCAAGAAAATGTACAACCTG.................  849

    262 rgLysAsnLeuLysSerLeuIleIleValHisProSer..........  274
                          |||:::::   ||||||
    850 ........          CCGATAGTAATGATGCCCTCTTTCGAAGTATTT 882

    275 TrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerValLy 291
        |||:::           |||   |||   :::      ||||'|
    883 TGGATAAAGATATAATTTTTGATCATATTAAAG.....ATTTCGAGCCA  925
```

```
     291 sPheIleAsnLysIleGlnTyrValHisSerIleGluAspLeuGlu..Gl 307
         :::   |:|||:::     |||      |: :::   ||||::  ||
     927 CTATTCGAACAAATTGTGGCGGGTA...CTGCTAAACATGTTACGGGACA 973

     307 nLeuIleProMetGluHisValGlnIleProAspCysValLeuGlnTyrG 324
         ' :::' ' :::   | :::'      :::     |
     974 AGAATTGCCAATGAAACCTGTACCATTACCA.........GTGGCAGAAG 1014

     324 luGluGluArgLeuLysAlaArgArgGluSerAlaArgProGlnProGlu 340
         ' :::::   || :::::  ::: |  |||: :
    1015 AGCCCGCTCAAGTACCAGCGGAAGAATTAGATCCACTCCAGAGGATGAC 1064

     341 PheValLeuProArgSerGluGluLysProGluValAlaProValGluAs 357
         |||:::|||        ||| :|     ||| ||      |||   ::
    1065 TTCGCATTAGATGTTACAGAATCTCCCGAGGAAGTAGAATTAGTATTAGA 1114

     357 nArgSerAlaLeuValSerGluAspGlnGlu 387
         :::::::|        :::|:::   |'
    1115 TGAAGAGGCAACTGAAGAAGAATCAACGGAA 1145

seq_[illegible] ...gra_P g'''9a'a/1 .ra/9R_DIME/seq.US-09-323-170B-1

seq_documentation_block:
; Sequence 1, Application US/08323170B
; Patent No. 5733772
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Kim C.
;    APPLICANT:  Keelow, David C.
;    TITLE OF INVENTION:  Cloning and Expression of Plasmodium
;    TITLE OF INVENTION:  falciparum Transmission-Blocking Target Antigen, Pfs230
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,170B
;      FILING DATE:  13-OCT-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/010,409
;      FILING DATE:  29-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Quine, Jonathan A.
;      REGISTRATION NUMBER:  P-41,261
;      REFERENCE/DOCKET NUMBER:  015288-[illegible]
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  149..9556
US-09-323-170B-1

alignment_scores:
            Quality:   87.00            Length:    384
```

```
            Ratio:  0.524               Gaps:      17
 Percent Similarity:  43.229   Percent Identity:  1.[illegible]

alignment_block:
 US-09-174-937-2 x US-08-323-170B-1

 Align seq 1/1  to: US-09-323-17[illegible] 1-1  from: 1  [illegible]

       3 ThrThrGluAlaThrLeuArgMetGluAsnValAspValLys[illegible]Tr 1+
         ||  :::::::|'|:::::::   ::: '
    1247 ACTTATGATAGTACAATAAAAAATGAAGATGTAGAT.........GAAGA 1+7

      19 pGlnAspGluAspLeuProArgProLeuProG.GluThrSerVa.S[illegible] ++
           |:|:::::::   :::  :::    :::
    1286 GGTAGGTGAAGAGGTAGGTGAAGAGGTAGGTGAAGACGTAGTTGAAGAA. 1+6

      36 euLeuGlySerProValGluAspThrSerSerProProAsn[illegible]As 5+
         :::  :::  '   :::
    1337 ..GTAGGTGAAGAGGTAGGTGAAGAG........................... 1++

      53 PheAsnGlyAlaHisArgLysArgLysThrLeuValAlaPr[illegible]As ++
                                          :::     :::::
    1361 ..............................GTAGGTGAAGAGGTAGG 1+77

      69 nIleSerLeuAspGlnSerGluGlySerLeuLeuSerAsp.........A ++
         :  ::::::  ::::::::|||  :::   :::::::::  :  :
    1378 TGAAGAGGTAGGTGAAGAAGAAGGTGAAGAGGTAGGTGAAGTAGTTGTAGGTG 142[illegible]

      83 spPheLeuAspThrProAspAspLeuAspIleLeuValAspAspIleGl[illegible] ++
         ::  :::       :::      :::::::::  ::::::
    1428 AACAGGTAGGTGAAGAAGAAGGTGAAGAGGTAGTTGTAGATGATGATGTTGTAA 147[illegible]

     100 ThrProAspGluThrAspSerLeu...GluPheLeuGlyAsnGlyAsnGl 11[illegible]
           ' |  '   :::          :::          :::
    1478 TATGTAGATGAAAAAGAAAGGCAAGTTGTAAATTTATCATTTGGTAATGA 152[illegible]

     115 uLeuGluTrpGluAspAspThrProValAlaThrAlaLysAsnMetPro[illegible] 1+2
         |   ||'|  ::::::        :    :  :::::
    1526 AGAAGAAAAAGATGAAGGTGGAGAAAGTTTTAT.TATGAAAATATTGAAA 157[illegible]

     132 lyAspSerAlaAspLeuPhe..............GlyAspSerTyrThr... 14[illegible]
           :::::::          :::          :::
    1578 TTGATAAAACAGATTTGTTTAATTTATATAATGAGTTAATGATGATGAT 162[illegible]

     143 ......ThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrVa 1[illegible]
               ' ''''' |
    1628 GTATATAAAGTGGATGGTTCC.......................AAAAT 167[illegible]

     157 lIleIleGlyGluGlnGluHisArgIleAspLeuHisMetIleArgProT 174
         |::::::   ::::::          :::    :
    1654 TTTATTAGATGATGATACA..................ATTATAGAGTAT 169[illegible]

     174 yrMetLysValValThrHisGlyGlyTyrTyrGlyGlu.............. 1[illegible]
           :::::::  :::::::::                 :
    1686 TTAAAAAACATATTGCATGATATTTCTAATATTTTCTAATATTTGAAGTT 173[illegible]

     187          GlyLeuAsnAlaIle[illegible]eValIleAlaAlaCysIle..... 1[illegible]
                  '''  :::   :::::::  :::  :::
    1736 GTGGAAGATGGAGAAAATGTTATAAAAATAATTAGAAGTTTTA AAAA 178[illegible]

     199 .......LeuProAspSerSerLeuProArgTyrHisTyrIleMetIleA 21[illegible]
                 ''''':::   ::::::  :::                :
    1786 TGGTGCATTACCAAGTGTAGGTGTTGATGATTAGATAAAATT... ...T 1[illegible]

     213 snLeuPheLeuTyrValIleSerSerLeuIleLeuValIleAlaIleAsp 22[illegible]
         ::|'          :::   :::   :::          :::
    1830 ATTTGTCATATCGAAACAACAGAAATCTCCACATATGTGTTATTTCAAATAT 187[illegible]

     230 ..........................................TyrMetIleVa 2[illegible]
                                                  :::
    1880 TCATATGATAAATATGCATCTAATAATACAAATAAATAACATTTTGTTA 192[illegible]

     233 lTyrLeuAsnGlyAlaThr[illegible]rMetProGlyIleGlyTrpL 2[illegible]
```

```
1930 TTTTACAGATCAATTAAAACCAACAGAAAGI33.CCIAAAGTA..... 1972

 250 euLysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnLeuLys 266

1973 ..AAAAAATGT                                      ..... 1981

 267 SerLeuIleIleValHisProSerTrpPheIleArgThrValLeuAlaIl 283

1982 .                .      ...      .....GAAGTAAAAGT 1992

 283 oSerArgProPheIleSerValLysPheIleAsnLysIleGln....TyrV 299

1993 TAATGAGCCATTAATAAAAGTAAAAATAATATGTCCATTAAAAGGTTCTG 2042

 299 alHisSerLeuGluAspLeuGluGlnLeuIleProMetGluHisValGln 315

2043 TAGAAAAATTATATGATAATATAGAATATGTAGTTAAAAAAAGC..... 2086

 316 IleProAspCysValLeuGlnTyrGluGluGlnArgLeuLysAlaArgAr 332

2087 ...CCATATGTTGTTTTAACAAAAGAGGAAACTAACTAAAGCAAAAACT 2133

 332 gGluSerAlaArgProGlnProGluPheValLeuProArgSerGluGluL 349

2134 TCTCTCGAAACTTATTTATGGTTTATTAATATCTCCGACGGTTAACGAAA 2183

 349 ys 349

2184 AG 2185
```

seq_name: /tmp/EFS/ptodata/1/ina/CC_COMP_seq_US-08-568-459A-3

```
seq_documentation_block:
; Sequence 3, Application US/08568459A
; Patent No. 5849306
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/568,459
;      FILING DATE:  07-DEC-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.3.1CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4507 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
US-08-568-459A-3


alignment_scores:
            Quality:   87.00            Length:    191
              Ratio:  0.956              Gaps:     10
 Percent Similarity:  47.644   Percent Identity: 25.654

alignment_block:
 US-09-174-937-2 x US-08-568-459A-3  ..

 Align seg 1/1  to: US-08-568-459A-3  from: 1  to: 4507

       4 ThrGluAlaThrLeuArgMetGluAsnValAspValLysGluGluTrpGl 23

    2961 ACTGAAGATGTTATTAGAAAAGAAAATCTTGACAAGCATGAAGATGAAAA 3010

      20 nAspGluAspLeuProArgProLeuProGluGluThrGlyValGluLeuL 67

    2911 AGGCGCAGATGAAGAAAGACATAGTACTTCTGAAAGC ..........T 2948

      37 euGlySerPro..................................... 46

    2949 TAAGTTCACCTGAAGAAAAAATGTTAACTGATAATGAAGGAGGAAATAGT 2998

      41 . .        .        ValGl.AspThrSerSerProPr:AsnThr.. 50

    2999 TTAAATCATGAAGAGGTGAAAGAACATACTAGTAATTCTGATAATGTTCA 3048

      51 . .        .     LeuAsnPheAspGlyAlaHisArgLysArgLysT 82

    3049 ACAGTCTGGAGGAATTGTTAATATGAATGTTGAGAAAGAACTAAAAGATA 3098

      62 hrLeuValAlaProGluIleAsnIleSerLeuAspGln.......... 74

    3099 CTTTAGAAAATCCT......TCTAGTAGCTTGGATGAAGGAAAAGCACAT 3142

      75          SerGluGlySerLeuLeuSerAsp . AspPheLeuAspTh 87

    3143 GAAGAATTATCAGAACCAAATCTAAGCAGTGACCAAGATATGTCTAATAC 3192

      87 rProAspAspLeuAspIleAsnValAspAspIleGluThrProAspGlnT 136

    3193 ACCTGGACCTTTGGAT   AACACCACTGAA   GAAACTACAGAAAGAA 3236

     104 hrAspSerLeuGluPheLeuGlyAspGlyAspGluLeuGluTrpGlu... 119

    3237 TTAGTAATAATGAATATAAAGTTAACCAGAGGAAGATGAGAGAACGCTT 3286

     120 ..        .. AspAspThrProValAlaThrAlaLysAsnMetProGl 132

    3287 ACTAAGGAATATGAAGATATTGTTTTCAAAAGTCATATGAATAGAGAATC 3336

     132 yAspSerAlaAspLeuPheGlyAspGlyThr      ............... 148

    3337 AGACGATGGTGAATTATATGACGAAAATTCAGACTTATCTACTGTAAATG 3386

     143 .....ThrGluAspGlySerAla 148

    3387 ATGAATCAGAAGACGCTGAAGCA 3409
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2000, 20:46:11 ; Search time 284.47 Seconds
(without alignments)
2608.604 Million cell updates/sec

Title: US-09-174-937-7
Perfect score: 2966
Sequence: 1 gtggaaaaatgcgatgtgggg[illegible] ... .aaaaaaaaaagggcggcccgc 2966

Scoring table: IDENTITY_NUC

Searched: 311595 seqs, 105096042 residues

Database : N_Geneseq_36:*

Word size : 0

Number of hits that pass the threshold . 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266.4	9.0	1227	1	X39873	Gastric cancer ass
2	255.4	8.6	816	1	X39875	Gastric cancer ass
c 3	75.2	2.5	19124	1	T72892	Plasmodium var-7 g
c 4	71.8	2.4	7797	1	X33180	Cowpox virus bsr f
c 5	71.8	2.4	6644	1	X33181	Base sequence of t
c 6	71.8	2.4	7372	1	X33182	Base sequence of t
c 7	71.8	2.4	7996	1	X33184	Base sequence of t
c 8	70.8	2.4	9789	1	T41852	cDNA encoding Plas
c 9	70.4	2.4	240	1	T76782	Staphylococcus aur
10	69.8	2.4	2104	1	Q225273	Sequence encoding
11	67.8	2.3	1100000	1	X20249_00	Continuation (7 of
12	67.8	2.3	1111000	1	X20250	Borrelia burgdorfe
c 13	67.4	2.3	2447	1	V54587	Human secretory pr
c 14	67.2	2.3	202	1	V41451	Nucleotide sequenc
c 15	64.4	2.2	32367	1	V35620	Human SHOX (short
c 16	64.4	2.2	15577	1	V35619	SHOX gene prelimin
c 17	62.4	2.1	2427	1	Q04107	Human pro-urokinas
c 18	61.8	2.1	4237	1	V61487	Human secreted pro
19	61	2.1	1627	1	Q47761	Antisense RNA comp
20	61	2.1	1100000	1	V21209_00	Continuation (9 of
21	60.8	2.0	2419	1	Q27999	P.falciparum GBP13
c 22	60	2.0	1671	1	Q24134	50 kD subunit of S
c 23	59.8	2.0	44453	1	X23519	Human kidney amino
c 24	59.4	2.0	5952	1	Q11710	Dictyostelium plas
25	58.8	2.0	4590	1	N60472	Sequence encoding
c 26	58.4	2.0	2503	1	Q53480	pNPX30 xylanase cD
c 27	58.2	2.0	2101	1	Q02047	Sequence encoding
c 28	58.2	2.0	4590	1	N60472	Sequence encoding
c 29	58	2.0	3035	1	Q03875	Sequence encoding
c 30	57.2	1.9	1100000	1	X20248_00	Continuation (3 of
c 31	56.6	1.9	2123	1	I59701	PTH-like peptide D
c 32	56.4	1.9	1189	1	V29221	Nucleotide sequenc
33	56.2	1.9	1402	1	V60073	Immunoreactive lam
34	55.8	1.9	2750	1	Q38808	G6PD coding sequen
c 35	55.8	1.9	51952	1	V26084	Tomato pest resist
36	55.6	1.9	19124	1	T72892	Plasmodium var-7 g
37	55.4	1.9	3994	1	X20261	Borrelia burgdorfe
c 38	55.2	1.9	1834	1	Q04690	Encodes Mammalian
c 39	55	1.9	605	1	T31530	Human 3' apolipopr
c 40	54.6	1.8	3711	1	V44271	Rabies virus Pm strain BM
41	54.4	1.8	3111	1	[illegible]	Sequence encoding
c 42	54.4	1.8	6124	1	[illegible]	Sequence encoding
c 43	54.2	1.8	2338	1	[illegible]	Plasmid pNX1 compr
44	54	1.8	4140	1	[illegible]	Aplysia californicum
c 45	53.8	1.8	5760	1	N71854	Sequence encoding

ALIGNMENTS

RESULT 1
X39873
ID X39873 standard; DNA; 1227 BP.
AC X39873;
DT 02-JUL-1999 (first entry)
DE Gastric cancer associated [illegible]
KW Cancer associated antigen; [illegible] therapy; [illegible] diagnosis; human;
KW breast cancer; colon cancer; [illegible] cancer; [illegible] cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; 98WO-CA00679.
PR 22-JUN-1998; 98US-102322.
PR 17-JUL-1997; 97US-896164.
PR 10-OCT-1997; 97US-061599.
PR 10-OCT-1997; 97US-061765.
PR 10-OCT-1997; 97US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, OHare M, [illegible] Y, [illegible] LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
DR WPI; 99-131446/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients useful in [illegible] for
PT for the diagnosis, monitoring or treatment of [illegible]
PS Claim 67; Page 579-580; 787pp; English.
CC The invention relates to a method for discovering a disorder characterised
CC by expression of a human cancer associated antigen, [illegible] for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product of the NAM or an expression
CC product complexed with an MHC class I molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The probe and NAM can be used in
CC the diagnosis, monitoring, prevention or treatment of cancers
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigens predicted expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, [illegible] cancer and
CC lung cancer.
SQ Sequence 1227 BP; 388 A; 270 C; 314 G; 254 T;

Query Match 9.0%; Score 266.4; DB 1; Length 1227;
Best Local Similarity 99.2%; Pred. No. 4.9e-47;
Matches 278; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 123 gactactgtgcgagggatgtaggg[illegible] 182
Db 880 GACTACTGTGCGAGGGATGTAGGGA[illegible] 939
Qy 183 gtgctgcagacatgaggcctgat[illegible] 242
Db 940 GTGCTGCAGACATGACGCCTGA[illegible] 999
Qy 243 ggagaatgaaggctgacagccgc[illegible] 302
Db 1000 GGAGAATGAAGCTGACAGCCGC[illegible] 1059
Qy 303 ttctctctgatgataac-ttgc[illegible] 361

```
Db   1060 TTCTCTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTT 1119

Qy    362 gataccccccgatgaagcagattcttttgagtacactggcc 401
          ||||||||||||||||||||||||||||||||||||||||||
Db   1120 GATACCCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCC 1159


RESULT   2
X39876
ID   X39876 standard; DNA; 816 BP.
AC   X39876;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer, colon cancer, gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; 98146?9.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061766.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 591, 797pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule, and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   816 BP;   231 A;   191 C;   209 G;   184 T;

  Query Match             9.0%;  Score 264.4;  DB 1;  Length 816;
  Best Local Similarity   99.3%;  Pred. No. 4.7e-40;
  Matches  278;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    123 gactactgtcagaggatgtaggaatggacatcgccttttgaagagggcgtgctgagtccca 182
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    469 GACTACTGTCAGAGGATGTAGGAATGGACATCGCCTTTTGAAGAGGGCGTGCTGAGTCCCA 528

Qy    183 gtgctgcagaacatgaggcctgaagcctcctaa?ctctggatcttaatgacactcatcctc 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    529 GTGCTGCAGACATGAGGCCTGAAGCCTCCTAA.?CTCTGGATCTTAATGACACTCATCCTC 588

Qy    243 ggaggaatcaagctcacagccccaaatatcaa?ctttctctggaccaaagtgaaggatcta 302
          |||||||||||||||||||||||||||||||?|||||||||||||||||||||||||||||
Db    589 GGAGGAATCAAGCTCACAGCCCCAAATATCAA?CTTTCTCTGGACCAAAGTGAAGGATCTA 648

Qy    303 ttctctctgatgataac-ttggacagtccagatgaaattgacatcaatgtggatgaacttt 361
          ||?|||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    649 TTTTTTCTGATGATAACTTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTT 708

Qy    362 gataccccccgatgaagcagattcttttgagtacactggcc 401
          ||||||||||||||||||||||||||||||||||||||||||
Db    709 GATACCCCCGATGAAGCAGATTCTTTTGAGTACCCTGGCC 748


RESULT   3
T72882/c
ID   T72882 standard; cDNA; 19124 BP.
AC   T72882;
DT   12-SEP-1997  (first entry)
DE   Plasmodium var-7 gene.
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   exon            7317..15139
FT                   /*tag= a
FT                   /number= 1
FT   intron          15140..16205
FT                   /*tag= b
FT                   /number= 1
FT   exon            16206..17552
FT                   /*tag= c
FT                   /number= 2
FT                   /note= "no stop codon given"
PN   WO9640766-A2.
PD   19-DEC-1996.
PF   07-JUN-1996; U09508.
PR   07-JUN-1995; US-487826.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI   Wellems TE;
DR   WPI; 97-052231/05.
DR   P-PSDB; W22475.
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
PS   Claim 4, Page 56-61, 96pp, English.
CC   This sequence represents the var 7 gene of Plasmodium. Var 7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72883 and T72889 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and
CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
SQ   Sequence   19124 BP;   7824 A;   2140 C;   2790 G;   6332 T;

  Query Match             2.5%;  Score 75.2;  DB 1;  Length 19124;
  Best Local Similarity   48.6%;  Pred. No. 2.1e-35;
  Matches  241;  Conservative    0;  Mismatches  248;  Indels    2;  Gaps    1;

Qy   2233 attttaaaatctaccttaatttt?ctttaaaatatatagtgtttttttttttttttt 2292
          || | |  || ||  |  | ||.  |  ||||.  |||||||||||||||||||||
Db  16185 ATATGTGTATGTATATATGTATTTAGGTATTATATATATGTGTTTCTTAATTTATATATTTA 16126

Qy   2293 aaaaaaaaagtgccttctgaata?ttttttttttaaaaagtgaaaaaatttttaaaaaaaa 2352
               | | |  | |  |||  ?| | | |  ||  | |  |  |  |  |  |  |
```

```
Db  16125 TGTAGTTATGGGTATATATATATATATATATAATATATATATGTATGTGTAAATGTGTAT 16066

Qy   2353 gtataagagggttggattataaatatgtttgagttctgaagtatgaaataaatgcttgaa 2412
          ||| |||| |||| | || | | || | | || |  |  ||  | |
Db  16065 GTAGTTATGGGTATATATATATATATATATATGTATGTGTAAATGTTGTATGTAGTTATGGG 16006

Qy   2413 agcagcctgtttcttaggtttctcaggcgaccattttgaaccaggaatgtacattctaaa 2472
          |  |  |  || || | |  |  |  |  | | ||  || | |
Db  16005 TATATATATATATATATGTATGTATATGTGCATGTAATTAATTTTATTTTATTTTTTTTA 15946

Qy   2473 caaaatagggtttttaaggtatagttttttcaaaaacatttgggcca·····taattt 2525
          |||  |||| | | | | || || | | |||  |  | |||| |
Db  15945 TTTAATTAAATTTTTTTATTTTTTTTTATTTCATTAAATTTTTTTAATTTTTTTTATTT 15886

Qy   2526 attttcttttaatgaatacttttttcatttatattatatttttcgtgtatatattttcatt 2585
          | ||  |||  | | |  || ||  ||| |||  |||  | | ||| | ||  ||
Db  15885 AATTAAATTTTTTTATTTATTATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15826

Qy   2586 atatagttttcattatatattatatatttgtactatatatttttcatttatatagttttc 2645
          | |  ||| | || || || | |||||  ||||||||||  ||||  ||||
Db  15825 TTTTTTATTTAAATAAATTTTTTTTATTTTATGTATATATTTTTTTTTTAACATTTTTTT 15766

Qy   2646 attgtatattatatacctcattatattatataattcattatatagttttcatttaaattta 2705
          | | | | ||  |  | ||  | | |  || | ||||| ||||| |||  |||
Db  15765 AATTTTTTTTTATTTTATGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTT 15706

Qy   2706 tttcattttataacta 2721
          ||  | | |||  ||
Db  15705 GTTTTTATGATATATA 15690
```

RESULT 4
X33180/c
ID X33180 standard; DNA; 7797 BP.
AC X33180;
DT 25-JUN-1999 (first entry)
DE Cowpox virus bsr full length gene sequence.
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Cowpox virus.
PN WO9913073-A2.
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DR WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 1; Page 34-38; 51pp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the cowpox virus bsr gene which
CC is used in an example from the present invention.
SQ Sequence 7797 BP; 2542 A; 1760 C; 1956 G; 1839 T;

```
  Query Match              2.4%;  Score 71.8;  DB 1;  Length 7797;
  Best Local Similarity   56.0%;  Prob. No. 7.8e-15;
  Matches  136;  Conservative    0;  Mismatches  ...;  Indels  ...;  Gaps    0;

Qy   2484 tttttaaggtatagtttttcaaaaacatttgggccatattttattttttttaattaatta 2543
          ||||  |  | | | | |
Db   5122 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5063

Qy   2544 cttttcatttatattatattttcatgtatatatttcattatatagttttcatttaaattta 2603
          ||||| || | | | | ||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   5062 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5003

Qy   2604 attatatatttgtactatatatttttcatttatatagtttt·attatattattatatt 2663
          ||  |  || |  | |
Db   5002 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4943

Qy   2664 cattatattatataattcattatatagtttcatttaaattt·····tttaaa 2723
          | | || | |  | | |  |  |  |  |  |  |  |  |  |  |  |  |
Db   4942 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTTTAA 4883

Qy   2724 aca 2726
          |||
Db   4882 AGA 4880
```

RESULT 5
X33181/c
ID X33181 standard; DNA; 6644 BP.
AC X33181;
DT 25-JUN-1999 (first entry)
DE Base sequence of the plasmid pRx-ires-bsr.
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Synthetic.
OS Cowpox virus.
PN WO9913073-A2.
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DR WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 1; Page 38-41; 51pp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC is used in an example from the present invention.
SQ Sequence 6644 BP; 2166 A; 1574 C; 1424 G; 1480 T;

```
  Query Match              2.4%;  Score 71.8;  DB 1;  Length 6644;
  Best Local Similarity   56.0%;  Prob. No. 7.8e-15;
```

```
  Matches  136;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    2484 tttttaaggtatagtttttcaaaaacatttgggccataatttatttttctttaatgaata 2543
           ||||| | | |||| |||          ||| |  ||| |||| |||| |   |
Db    3969 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTT 3910

Qy    2544 ctttttcatttatattatttttcgtgtatatattttcattatatacgtttcattatat 2603
           ||||| ||| | || | ||||| | |  | |||  || | | |  ||| || | |
Db    3909 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3850

Qy    2604 attatatatttgtactatatatttttcatttatatagttttcattgtatattatatacct 2663
           || | | ||| | | | | ||||| ||| | | ||||  || | | || | | |   |
Db    3849 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3790

Qy    2664 cattatattatataattcattatatagttttcatttaatttatttcatttttataactaga 2723
           || | || | | |  || | | |  |||| |||  ||| ||| ||||  |||| |
Db    3789 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAACTGCA 3730

Qy    2724 aca 2726
           | |
Db    3729 AGA 3727


RESULT   6
X33182/c
ID   X33182 standard; DNA; 7372 BP.
AC   X33182;
DT   25-JUN-1999  (first entry)
DE   Base sequence of the plasmid pRx Bcl-xl-bsr.
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9913073-A2.
PD   18-MAR-1999.
PF   07-SEP-1998; J04010.
PF   08-SEP-1997; JP-259235.
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI   Hamada H;
DR   WPI; 99-243729/20.
PT   New apoptosis-resistant virus-sensitive cell
PS   Example 2; Page 41-45; 51pp; English.
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and
CC   is used in an example from the present invention.
SQ   Sequence   7372 BP;    2353 A;    1719 C;    1649 G;    1621 T;


  Query Match             2.4%;  Score 71.8;  DB 1;  Length 7372;
  Best Local Similarity   56.0%;  Pred. No. 7.8e-05;
  Matches  136;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    2484 tttttaaggtatagtttttcaaaaacatttgggccataatttatttttctttaatgaata 2543
           ||||| | | |||| |||          ||| |  ||| |||| |||| |   |
Db    4697 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4638

Qy    2544 ctttttcatttatattatttttcgtgtatatattttcattatatacgtttcattatat 2603
           ||||| ||| | || | ||||| | |  | |||  || | | |  ||| || | |
Db    4637 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4578

Qy    2604 attatatatttgtactatatatttttcatttatatagttttcattgtatattatatacct 2663
           || | | ||| | | | | ||||| ||| | | ||||  || | | || | | |   |
Db    4577 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4518

Qy    2664 cattatattatataattcattatatagttttcatttaatttatttcatttttataactaga 2723
           || | || | | |  || | | |  |||| |||  ||| ||| ||||  |||| |
Db    4517 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAACTGCA 4458

Qy    2724 aca 2726
           | |
Db    4457 AGA 4455


RESULT   7
X33184/c
ID   X33184 standard; DNA; 7996 BP.
AC   X33184;
DT   25-JUN-1999  (first entry)
DE   Base sequence of the plasmid pRx-Bcl-2-i-h2D-25.
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9913073-A2.
PD   18-MAR-1999.
PF   07-SEP-1998; J04010.
PR   08-SEP-1997; JP-259235.
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI   Hamada H;
DR   WPI; 99-243729/20.
PT   New apoptosis-resistant virus-sensitive cell
PS   Example 3; Page 46-49; 51pp; English.
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-Bcl-2-i-h2D-25, which contains the human Bcl-2 gene, and
CC   is used in an example from the present invention.
SQ   Sequence   7996 BP;    2463 A;    2015 C;    1829 G;    1689 T;


  Query Match             2.4%;  Score 71.8;  DB 1;  Length 7996;
  Best Local Similarity   56.0%;  Pred. No. 7.9e-05;
  Matches  136;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    2484 tttttaaggtatagtttttcaaaaacatttgggccataatttatttttctttaatgaata 2543
           ||||| | | |||| |||          ||| |  ||| |||| |||| |   |
Db    5321 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5262
```

```
Qy   2544 ctttttcatttatattatatttttgctgtatatatttttcattatatatgtttcattatat 2603
          ||| |   | | || | |||||  | | |   ||||  || | |   |||  || |
Db   5261 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5202

Qy   2604 attatatatttgtactatatatttttcatttatatagttttcattgtatattatatacct 2663
          | | | | | |   |  |||| | |  | ||||  || | | || | |   |  |
Db   5201 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5142

Qy   2664 cattatattatataatttcattatatagttttcatttaatttatttcattttataactaga 2723
          || | || | |  ||  | | | |  |||| |||  ||| |||  ||||  ||||  |
Db   5141 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAACTGCA 5082

Qy   2724 aca 2726
          | |
Db   5081 AGA 5079


RESULT   8
T41852/c
ID   T41852 standard; DNA; 9789 BP.
AC   T41852;
DT   20-FEB-1997  (first entry)
DE   cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
KW   Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW   detection, identification, treatment, prevention, parasite, ss.
OS   Plasmodium falciparum MC type.
FH   Key             Location/Qualifiers
FT   cds             326. .9497
FT                   /*tag= a
FT                   /product= Erythrocyte membrane protein
FT   misc_feature    518. .520
FT                   /*tag= b
FT                   /transl_except= GTA encodes Tyrosine
FT   misc_feature    656. .658
FT                   /*tag= c
FT                   /transl_except= ATT encodes Leucine
FT   misc_feature    2909. .2911
FT                   /*tag= d
FT                   /transl_except= AAC encodes Aspartic acid
FT   misc_feature    3461. .3463
FT                   /*tag= e
FT                   /transl_except= GAA encodes Glutamine
FT   misc_feature    5546. .5548
FT                   /*tag= f
FT                   /transl_except= CCT encodes Arginine
FT   misc_feature    6254. .6256
FT                   /*tag= g
FT                   /transl_except= AAT encodes Lysine
FT   misc_feature    6257. .6259
FT                   /*tag= h
FT                   /transl_except= ATA encodes Tyrosine
FT   misc_feature    6263. .6265
FT                   /*tag= i
FT                   /transl_except= AAC encodes Lysine
FT   misc_feature    6269. .6271
FT                   /*tag= j
FT                   /transl_except= TTC encodes Isoleucine
FT   misc_feature    6272. .6274
FT                   /*tag= k
FT                   /transl_except= ATA encodes Histidine
FT   misc_feature    6275. .6277
FT                   /*tag= l
FT                   /transl_except= ATT encodes Asparagine
FT   misc_feature    6278. .6280
FT                   /*tag= m
FT                   /transl_except= GGA encodes Tryptophan
FT   intron          7754. .8478
FT                   /*tag= n
PN   WO9638736-A1
PD   31-OCT-1996.
PF   26-APR-1996; WO05798.
PR   27-APR-1995; US-430908.
PA   (APPY-) APPYMAX TECHNOLOGIES NV.
PI   Baruch DI, Howard RJ, Pasloske BL;
DR   WPI; 96-497376/49.
DR   P-PSDB; W00384.
PT   New Plasmodium falciparum erythrocyte membrane protein, useful to
PT   develop products for the diagnosis, treatment and prevention of
PT   malaria parasite infection
PS   Disclosure; Figure 11; 149pp; English.
CC   A polypeptide comprising a Plasmodium falciparum (P.f.) erythrocyte
CC   membrane protein 1 (PfEMP1), or a part fragment, variant or mutant
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection. The polypeptides may inhibit binding or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 may be used as DNA
CC   probes and primers to identify a Plasmodium falciparum parasite. The
CC   primers used to generate characteristic amplification patterns from
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes the
CC   PfEMP1 protein of the MC type of Plasmodium falciparum. An
CC   alternative, truncated version of the coding sequence (cDNA clone)
CC   is given in T41853.
SQ   Sequence   9789 BP;    4061 A;    1344 C;    1447 G;    2937 T;

  Query Match              2.4%;  Score 7.8;  DB 2;  Length 9789;
  Best Local Similarity   37.1%;  Pred. No. 0.0012;
  Matches  129;  Conservative      0;  Mismatches  217;  Indels     0;  Gaps     0;

Qy   2493 tatagtttttcaaaaacatttg                                      2552
          |||| | | | |  |||||                                        |||
Db   8353 TATAATATATATATATATTATA                                     8294

Qy   2553 ttatattatatttttcgtgtat                                     2612
          |  |  |  |  |  |  |||                                     |||
Db   8293 TTTTTTTTTTTTTAATTTTTTA                                     8234

Qy   2613 ttgtactatatatttttcatttatttttttattattttttttttttttttttttttt 2672
          |  | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   8233 TTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 8174

Qy   2673 atataattcattatatagtttt                                     2732
          |  |  |  |  |  |  |||                                     |||
Db   8173 ATATTTTTTTTTAACATTTTTT                                     8114

RESULT   9
I76782/c
ID   I76782 standard; DNA; 240 BP.
AC   I76782;
DT   15-SEP-1997  (first entry)
DE   Staphylococcus aureus exfoliation toxin A gene partial cds.
KW   Asymmetric polymerase chain reaction; ds DNA; ss DNA amplification;
KW   PCR; detection; assay; exfoliation toxin A; ETA; skin infection;
KW   competitive primer; capture probe; ss.
OS   Synthetic.
PN   US5627054-A.
PD   06-MAY-1997.
PF   05-APR-1996; 628417.
PR   05-APR-1996; US-628417.
PA   (USSA ) US SEC OF ARMY.
PI   Gillespie D;
DR   WPI; 97-271311/24.
PT   Quantitative nucleic acid amplification using competitive primers
PT   asymmetric polymerase chain reaction
PS   Example 1; Column 6; 9pp; English.
CC   In a specific example of a novel process for amplifying DNA or RNA
CC   (known or unknown) of a double-stranded nucleic acid sequence to produce
CC   single-stranded nucleic acid in an amount that is proportional to the
CC   starting amount of the nucleic acid, the Staphylococcus aureus
CC   exfoliative toxin A (ETA) gene was used as the DNA template. The
CC   region comprising nucleotides 107-437 was amplified by asymmetric
CC   asymmetric or competitor primer asymmetric PCR using the primers
CC   ETA-A2 and ETA-B (see I76779 and I76780). The asymmetric PCR, the
```

```
CC   amount of primer ETA-B was reduced and for competitor primer
CC   asymmetric PCR a competitor primer ETA-B2 (see T76780) was added
CC   with upstream primer ETA-A2 after the initial cycling reaction. PCR
CC   products containing ETA-specific sequences were detected
CC   radioactively by a capture system which employed a bifunctional
CC   capture probe ETA CP (see T76781 and T76782). ETA-CP was designed
CC   to capture the amplified sense strand onto capture membranes
CC   through hybridisation between the first 40 nucleotides of ETA-CP
CC   and nucleotides 221-260 of the ETA gene and through hybridisation
CC   of the poly(dA) tail on ETA-CP with poly(dT) tails on the capture
CC   membranes. A radioactively labelled "label probe" (see T76783),
CC   complementary to nucleotides 389-410 of the ETA gene was used to
CC   detect the amplicons. Results showed that hybridisation of the
CC   capture probe and label probe to the denatured symmetric PCR
CC   product was much less efficient than hybridisation to the
CC   single-stranded PCR products of the asymmetric and competitor
CC   primer asymmetric reactions.
SQ   Sequence   240 BP;    221 A;    3 C;    4 G;    12 T;

  Query Match             2.4%;  Score 73.4;  DB 1;  Length 240;
  Best Local Similarity   58.7%;  Pred. No. 9.8e-05;
  Matches  122;  Conservative    0;  Mismatches   86;  Indels    2;  Gaps    2;

Qy    2523 tttatttttttttaatgaatacttttttcatttatattatatttttcgtgtatatatttttc 2582
           ||| ||||| ||||  |   |  |||||  ||  | || | |||||  | | | | ||||
Db     237 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 178

Qy    2583 attatatatagtttcattatatattatatatttgtactatatatttttcatttatatagtt 2642
           | | | |  ||| || | | || | | | | | ||||  ||| | |  ||
Db     177 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 118

Qy    2643 ttcattgtatattatatacctcattatattatataattcattatatagttttcatttaat 2702
           || || | || | |  | || | ||  |  || || | |  ||||  |||  |
Db     117 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 58

Qy    2703 ttatttcatttttataactgaaacaatga 2730
           | | |  ||| |  || | ||||
Db      57 TTTTTTTTTTTTTTTTTTAGTAAAATTA 30


RESULT  10
Q25273
ID   Q25273 standard; cDNA; 2104 BP.
AC   Q25273;
DT   22-NOV-1992  (first entry)
DE   Sequence encoding the Histidine-rich protein (HisRP) associated with
DE   the knob (K+) phenotype.
KW   Malaria vaccine; Histidine-rich protein; cytoadherence; ss.
OS   Plasmodium lophurae.
FH   Key             Location/Qualifiers
FT   cds             638..1847
FT                   /*tag= a
PN   US5116965-A.
PD   25-MAY-1992.
PF   26-AUG-1986; 900401.
PR   26-AUG-1986; US-900401.
PA   (SLOK ) SLOAN KETTERING INST CANCER.
PI   Pologe L. Parvatch JV.
DR   WPI; 92-199590/24.
DR   P-PSDB; R24392.
PT   Histidine-rich protein associated with Plasmodium knob phenotype -
PT   and DNA encoding it, used for in vitro diagnosis of P.
PT   Falciparum infection.
PS   Claim 1; Fig 4A-D; 29pp; English.
CC   Two variants of HisRP are produced by P. falciparum.  One is
CC   associated with what is referred to as "knobby phenotype" (K3O) and
CC   "knobless phenotype" (K-).  The "knobby" and "knobless" phenotypes
CC   have been implicated in cytoadherence, which is characteristic of
CC   erythrocyte infection.  It has now been found that cDNA expressing
CC   both K- and K+ HisRP can be obtained by the use of P. lophurae HisRP
CC   expressing DNA.
SQ   Sequence  2104 BP;    969 A;    304 C;    293 C;    638 T;

  Query Match             2.4%;  Score 69.8;  DB 1;  Length 21.4;
  Best Local Similarity   52.2%;  Pred. No. 0.00016;
  Matches  178;  Conservative    0;  Mismatches  162;  Indels    1;  Gaps    1;

Qy    2428 aggtttctcacgcgaccatttgaaaaaaaaatgtacattctaaaaaaaatttttttt 2487
           || ||    |   || |||||  |  | ||  | ||  ||  |  | | | | | |
Db     207 AGATTGAAAAATAGAAAATTTTATATAATAATAAAAAATTAAATTCATTTAATAAAAAT 266

Qy    2488 taaggtatagttttttcaaaaacatttgggcgataatttatttttttttaatttatttttt 2547
           | |   ||  |||||   | | |     | |||  ||  |  | | | | | | | |
Db     267 TTATAAATTATTTTTTTTATATAAATTTTTTTTTTTTTTTTTTAATTTATATTATTTTTT 326

Qy    2548 ttcatttatattatatttttcgtgtatatatttttcattatatacgtttcatttatttttt 2607
           ||  || || | | |          |  |  | | |||| | | | | | | | | | |
Db     327 AGAATACATTTTTTGTATAAAAATTTTTTCATATATATATACATATTTTGTTTTATTATA 386

Qy    2608 tatatttgtactatatatttttcatttatatattttcattatatattatatacgtttatt 2667
           | |||      ||||  |  || ||| || | ||  |  | |||
Db     387 ATTTTTTTATAATTATACAATATATATTTATTATTTTTTATTTTTTTTTTTTTTTTTTTT 446

Qy    2668 atattatataattcattatatatttttttttttttttttttttttttttttttttttttt 2727
           | || |||| ||||||||    || |  |  ||| |||  |||| |  |  |  |  |
Db     446 TTTTTTTATATTTCATTATTATATTATATTGCTTTTTTTTTTTTTTTTTTTTTTATAATA 505

Qy    2728 tgaaacacagagaagttaagtaatttacttaaagtcacaca 2768
           |  | || | |  |||||  |        ||| |  | | |
Db     506 TATAACATATATTAGTTATTTTTTACACCTAAAGAAAAATA 546


RESULT  11
X20248_06
Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi  polynucleoti
WP Sequence split into 10 fragments  EMBOS X20248 Accession XX 24+
WP     Fragment Name      Begin     End
WP     X20248_00              1    100000
WP     X20248_01         100001    200000
WP     X20248_02         200001    300000
WP     X20248_03         300001    400000
WP     X20248_04         400001    500000
WP     X20248_05         500001    600000
WP     X20248_06         600001    700000
WP     X20248_07         700001    800000
WP     X20248_08         800001    900000
WP     X20248_09         900001    910715

  Query Match             2.3%;  Score 67.8;  DB 1;  Length 111309;
  Best Local Similarity   58.2%;  Pred. No. 0.00055;
  Matches  138;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    1;

Qy    2523 tttatttttctttaatgaatacttttttca--tttatattatatttttcgtgtatatatttt 2580
           ||| | |  ||| ||| ||| |||  |  ||| | | | | | | | | | | | | |
Db  92642 TTTTTGTGATTTCAATAAATTCTTCAGCAATTTTTTTTAAAATGTTTTCTTTTTTTTTTTT 92701

Qy    2581 tcattatatacgtttcattatatttatatttgtactatatatttttcatttatttttttt 2640
           | || | ||   |||| || |  | | | |  | |  |  |  |  |  |  |  |  |
Db  92702 TAATAACATTGATTTCGTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 92761

Qy    2641 ttttcattgtatattatatacctcattatattatataattcattatatagttttcatttaat 2700
           |   ||| | | || |||  ||  | | | | | | | | | | | | | | | | |
Db  92762 TATATTATTTCTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 92821

Qy    2701 atttatttcatttttataactgaaacaatgaaaaacacacacacaaaatttaattaatttttt 2760
           | | ||  ||| | |||| |          ||  |  |  |  |  |  |  |  |  |
Db  92822 TTATTTTAAATTCTTTAACAATTATATTCTTTAACTTTTTATTTAATGAAAAATTTAT 92878


RESULT  12
X20250
ID   X20250 standard; DNA; 111309 BP.
AC   X20250;
```

```
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #3.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever, endemic relapsing fever, Lyme borreliosis,
KW   infection, diagnosis, characterisation, detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858948-A1.
PD   30-DEC-1998
PF   18-JUN-1999; 99WO-012764.
PR   03-SEP-1997; 97US-057483.
PR   20-JUN-1997; 97US-050359.
PR   22-JUL-1997; 97US-053344.
PR   22-JUL-1997; 97US-053377.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Iarbigra F, Smith HO,
PI   White OR;
DR   WPI; 99-061217/07.
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 729-800; 1120pp; English.
CC   X20248 to X21402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence  111309 BP;  30956 A;  18151 C;  19075 G;  43117 T;
```

```
  Query Match             2.3%;  Score 67.8;  DB 1:  Length 111309;
  Best Local Similarity   58.2%;  Pred. No. 0.000055;
  Matches  138;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    1;

Qy    2523 tttatttttttttaatgaatacttttttca--tttatattatatttttcgtgtatatattt 2580
           |  | |  |  | ||| | | ||  |      | | | | || ||| | | | | | | |
Db  103256 TTTTTGTGATTTCAATAAATTCTTCAGCAATTTTTTTTAAAATGTTTTCTTTGTTTTTTG 103315

Qy    2581 tcattatatacgtttcattatatattatatatttgtactatatatttttcatttatataa 2640
           |  |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  103316 TAATAACATTGATTTCGTTTTTTATTGTATTTTTTTGTATATTTTTTCTTTTTTGCTTG 103375

Qy    2641 ttttcattgtatattatatacctcattatattatataattcattatatagttttcattta 2700
           |  ||| | | | || ||| |  || ||| |  | |  |  || | | | |||||||| |
Db  103376 TAGATATTATTTCTTTTATTTCTTCTTTTATGTTTTTTATATTTGTTTTGATTTCATTTA 103435

Qy    2701 atttatttcatttttataactagaacaatgaatgcacagagaaagtttaagtaatttgcct 2757
           | | | |  ||  | |||| |  ||| | |    | |     | ||||  ||  | | |
Db  103436 TTATTTTAAATTCTTTAACAATTACATTGTTTAACTTTTTATTTAATGAAAATTCAT 103492
```

```
RESULT  13
V54587/c
ID   V54587 standard; cDNA; 2447 BP.
AC   V54587;
DT   30-OCT-1998  (first entry)
DE   Human secretory protein encoding cDNA clone CO1020-1.
KW   Secretory protein; AM-373-1; BK250-1; BS330-1; DI539-3; IN729-3;
KW   CO1390-3; CO1020-1; DS752-3; DM340-1; IW902-1; nutritional source;
KW   cytokine; cell proliferation; immune stimulation; immune suppression;
KW   haematopoiesis regulation; receptor/ligand activity; tumour inhibition;
KW   anti-inflammatory activity; cadherin; tumour invasion suppression;
KW   activin; inhibin; chemostatic; chemokinetic; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             184. .1191
FT                   /*tag= a
FT   CDS             347. .592
FT                   /*tag= b
```

```
FT                   /note= "ORF [illegible] this Bb [illegible] W75-41"
PN   WO9833916-A2.
PD   06-AUG-1998.
PF   30-JAN-1998; 98WO-GB00811.
PR   26-JAN-1998; 98US-014969.
PR   31-JAN-1997; 97US-792511.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merriam J,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-407469/37.
DR   P-PSDB; W75857, W75861.
FT   New nucleic acid encoding secreted protein from human [illegible] and
PT   related vectors, transformed cells and promoter
PS   Claim 30; Pages 80-82; 114pp; English.
CC   Sequences shown in V54581 to V54588 represent cDNA sequences encoding
CC   human secretory proteins (W75857, W75860). The sequences correspond to
CC   the clones designated AM-373-1, BK250-1, BS330-1, DI539-3, IN729-3,
CC   CO1390-3, CO1020-1, DS752-3, DM340-1 and IW902-1 respectively. The
CC   polynucleotides and polypeptides can be used as nutritional sources
CC   or supplements. They may exhibit [illegible] activity,
CC   differentiation activity, [illegible] stimulating or suppressing activity,
CC   haematopoiesis regulating activity, [illegible] activity, anti-
CC   inflammatory activity, [illegible] activity, tumour inhibition activity.
CC   They may also exhibit [illegible] such as tissue growth activity,
CC   [illegible] or chemokinetic activity.
SQ   Sequence  2447 BP;  802 A;  478 C;  [illegible]
```

```
  Query Match             2.3%;  Score 67.4;  DB 1:  Length 2447;
  Best Local Similarity   57.5%;  Pred. No. 0.0044;
  Matches  119;  Conservative    1;  Mismatches   87;  Indels    2;  Gaps    1;

Qy    2523 tttatttttttttaatgaatacttttttca--tttatattatatttttcgtgtatatattt 2580
           |  | |  |  | ||| | | ||  |      | | | | || ||| | | | | | | |
Db    2440 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2381

Qy    2581 tcattatatacgtttcattatatattatatatttgtactatatatttttcatttatataa 2640
           |  |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    2380 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2321

Qy    2641 ttttcattgtatattatatacctcattatattatataattcattatatagttttcattta 2700
           |  ||| | | | || ||| |  || ||| |  | |  |  || | | | |||||||| |
Db    2320 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2261

Qy    2701 atttatttcatttttataactagaacaatgaatgcacagagaaagtttaagtaatttgcct 2757
           | | | |  ||  | |||| |  ||| | |    | |     | ||||  ||  | | |
Db    2260 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2204
```

```
RESULT  14
V41451/c
ID   V41451 standard; cDNA; 202 BP.
AC   V41451;
DT   12-OCT-1998  (first entry)
DE   Nucleotide sequence of the 5' portion of the Bl229_22 cDNA.
KW   Clone Bl229_22; secreted protein; human cell; [illegible] FIR;
KW   anti-tumour agent; tissue [illegible] stimulant; [illegible] agent; primer;
KW   thrombolytic agent; [illegible] antibody; ds.
OS   Homo sapiens.
PN   WO9821330-A2.
PD   22-MAY-1998.
PF   14-NOV-1997; 97WO-US20740.
PR   13-NOV-1997; 97US-969515.
PR   10-NOV-1996; 96US-749745.
PR   02-JUN-1997; 97US-867678.
PA   (GEMY ) GENETICS INST INC.
PI   Agostini MJ, Jacobs K, Lavallie ER, McCoy JM, Merriam J,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-287996/26.
FT   New nucleic acid encoding secreted protein from human cells -
PT   potentially useful, e.g. as [illegible] [illegible] [illegible]
```

```
PT   promoters of tissue growth, haemostatic and thrombolytic agents
PS   Claim 33; Page 72; 98pp; English.
CC   This is the nucleotide sequence of the clone BL229_22, used in the
CC   method of the invention involving secreted proteins from human cells
CC   potentially useful as immunomodulators, anti-tumour agents, tissue
CC   growth promoters, haemostatic and thrombolytic agents.  It was isolated
CC   from an adult human testes cDNA library.  They may be used to express
CC   the recombinant protein as tissue/molecular weight markers, for
CC   chromosome identification, to identify possible genetic disorders and
CC   isolate new related DNA, as sources of primers for PCR, to generate
CC   anti-protein or anti-DNA antibodies, or in interaction trap assays to
CC   identify sequences that encode interacting proteins.
SQ   Sequence   202 BP;    199 A;    0 C;    0 G;    1 T;

  Query Match              2.3%;  Score 87.2;  DB 1;  Length 202;
  Best Local Similarity   58.8%;  Pred. No. 0.00036;
  Matches  114;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy   2522 tttatttttttttaatgaatactttttcatttatattatatttttcgtgtatatatttttc 2582
Db    197 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTN 138

Qy   2583 attatatagttttcattatatattatatatttgtactatatttttcatttatatagtt 2642
Db    137 ANTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 78

Qy   2643 ttcattgtatattatatacctcattatattatataattcattatatagttttcatttaat 2702
Db     77 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 18

Qy   2703 ttatttcatttttat 2716
Db     17 TTTTTTTTTTTTTT 4


RESULT  15
V35620/c
ID   V35620 standard; DNA; 32367 BP.
AC   V35620;
DT   07-SEP-1998  (first entry)
DE   Human SHOX (short stature homeobox containing gene) gene sequence.
KW   Homeobox domain; human growth gene; growth regulation; growth defect;
KW   turner's syndrome; short stature homeobox containing gene; SHOXa;
KW   SHOX; bone disease; osteoporosis; calcium regulation; short stature;
KW   transcription factor A; ss.
OS   Homo sapiens.
PN   WO9814568-A1.
PD   09-APR-1998.
PF   29-SEP-1997; E05355.
PR   16-JAN-1997; EP-100583.
PR   01-OCT-1996; US-027633.
PA   (RAPP/) RAPPOLD-HOERBRAND G.
PI   Rao E, Rappold-hoerbrand G;
DR   WPI; 98-271713/24.
PT   New human growth genes - used to develop products for the diagnosis
PT   and treatment of human growth defects such as short stature, e.g.
PT   Turner's syndrome
PS   Claim 19; Pages 51-67; 84pp; English.
CC   This is the human SHOX gene sequence containing the PAR1 region. The gene
CC   region corresponding to short stature has been identified as a region of
CC   approximately 500 kb in the PAR1 region of the X and Y chromosomes. Three
CC   genes in this region have been identified as candidates for the short
CC   stature gene. These genes were designated SHOX (also referred to as
CC   SHOX93 or HOX93), pET92 and SHOT (SHOX-like homeobox gene on chromosome
CC   three). The SHOX gene has two separate splicing sites resulting in two
CC   variations SHOXa and SHOXb. The specification provides sequences of SHOX
CC   (short stature homeobox containing) genes SHOX ET92, SHOXa, SHOXb, SHOT
CC   and exons of the SHOX genes as shown in V35610 to V35621 and protein
CC   sequences of the human growth protein transcription factor SHOXa, SHOXb
CC   and SHOT as shown W60573 to W60575. The novel genes are responsible for
CC   human growth. Defects in the genes can cause short stature, e.g.
CC   Turner's syndrome. The products can be used to develop agents for the
CC   treatment of short stature or other human growth disorders. The products
CC   can also be used for providing a stimulatory effect on cells, e.g. for the
CC   treatment of bone diseases such as osteoporosis and diseases involved
CC   with disturbance in the bone calcium regulation.
SQ   Sequence   32367 BP;    7627 A;    8140 C;    8564 G;    7 T;    0 other;

  Query Match              2.2%;  Score 84.4;  DB 1;  Length 32367;
  Best Local Similarity   46.6%;  Pred. No. 0.002;
  Matches  206;  Conservative    0;  Mismatches  236;  Indels    6;  Gaps    0;

Qy   2317 tatattttgtaaacatgaggacatgcataagtcactgtataaaaaaaaaaaaaaaaaaaa 2376
Db   7164 TTTTATATGTATATATATGTTAATATATATATATTTTTATATGTATATATTTTTATATATA 7105

Qy   2377 atgtttgagttctgaagtatgaaaaaaatctgaaaagaaaaaaaaaaaaaaaaaaaaaaaa 2436
Db   7104 ATTTTATGTATATATATGTTTATATATGTATATTTATATGTATATATGTTTATATATATATT 7045

Qy   2437 acgcgaccattttgaaccaggaatgaaattctaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2496
Db   7044 TCTTTATATGTATATATGTTTATATATATTTTTATATGTATATATATGTTTTTATATATA 6985

Qy   2497 gtttttcaaaaacatttgggccatgaaaaaattttcttttaaaaaaaaaaaaaaaaaaaaa 2556
Db   6984 TTTATATATATACAATATGTTTGTATATATTTATATATGTTTATATATTTATATGTTTAT 6925

Qy   2557 attatatttttcgtgtatatatttttcattatatacgttttcattatatatatatatttttt 2616
Db   6924 ATATTTATATATGTTTATATGTTTATATATATTTATACATATTTTATATATTTTAATATA 6865

Qy   2617 attatatatttttcatttatatagttttcattatatattatatatttattatatatatat 2676
Db   6864 TTTATATATTATATATGTTTATATATTTATGTTTATATATTATATATGTTTATATATTTA 6805

Qy   2677 aattcattatatagttttcatttaatttatttcatttttataactataaaaaaaaaaaaaaa 2736
Db   6804 TATATTATATATGTGTTTATATATTTATATATGTATATAJATTATATATATTTATATA'A 6745

Qy   2737 gagaagttaagtaatttgccta 2758
Db   6744 TTTATGTATGTTTATATATTTA 6723


Search completed: February  8, 2000, 20:54:01
Job time: 1489 sec
```

```
SM of: US-09-174-937-2 to. EST.*   Out_Format : pfs

Date: Feb 10, 2000  1:33 PM

About: Results were produced by the GenCore software, version 4.5.
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV-x1p
 -Q=/cgn2_1/USPTO_app0_1/US09174937/runat_1_222203_995107_18273/app_query.fasta_1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPOL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09174937
 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-174-937-2
Query length: 371
Database: EST:*
Database sequences: 4038634
Database length: 1887831982
Search time (sec): 1131.150000



score_list:
Sequence              Strd Orig    ZScore    EScore Len   ' Documentation ...
gb_est27:AI437189       -  595.50 1250.32   7.1e-61   499  ' AI437189 fb38g06.y1 Zebrafish W
gb_est1:T05645          -  592.00 1246.51   1.0e-60   368  ' T05645 EST03534 Fetal brain, St
gb_est1:T31973          +  496.50 1040.88   9.9e-49   423  ' T31973 EST41767 Human Uterus Ho
gb_est1:T08680          +  449.00  939.05   1.3e-43   437  ' T08680 EST06572 Infant Brain, H
gb_est10:AA137873       -  426.00  890.67   7.6e-41   449  ' AA137873 mq77d03.r1 Stratagene
gb_est1:232819          -  422.00  885.44   1.7e-40   312  ' 232819 HHEA11F Atrium cDNA lib
gb_est14:AA450831       +  419.00  877.05   4.4e-40   368  ' AA450831 vq55e03.r1 Beddington
gb_est9:AA067188        -  402.00  838.27   6.3e-39   402  ' AA067188 mm39b12.r1 Stratagene
gb_est9:AA068579        +  402.00  838.51   6.1e-38   443  ' AA068579 mm45e06.r1 Stratagene
gb_est21:AA914577       -  392.00  815.32   1.2e-36   516  ' AA914577 vq02e07.r1 Soares mous
gb_est9:AA113114        -  390.00  814.66   1.3e-36   382  ' AA113114 mn68h02.r1 Stratagene
gb_est29:AI620433       -  368.00  764.78   7.3e-34   486  ' AI620433 tu47b05.x1 NCI_CGAP_Pr
gb_est1:T31983          -  357.00  747.98   6.8e-33   281  ' T31983 EST41823 Human Uterus Ho
gb_est10:AA129380       -  342.00  711.78   7.0e-31   394  ' AA129380 zn84e05.r1 Stratagene
gb_est32:AI722140       -  326.50  679.54   9.5e-29   604  ' AI722140 fd19g03.y1 Suzano Kawa
gb_est27:AI410693       -  317.50  657.57   7.7e-28   401  ' AI410693 EST240933 Normalized
gb_est1:T32180          -  310.00  652.47   1.1e-27   186  ' T32180 EST44614 Human Brain Hom
gb_est2:T94657          -  304.00  629.27   2.9e-26   441  ' T94657 yo37c09.r1 Stratagene Li
gb_est24:AU035341       -  300.00  619.41   1.1e-25   910  ' AU035341 AU035341 Sugano mouse
gb_est34:AI798173       -  294.00  605.61   5.7e-25   594  ' AI798173 qi18c09.y1 Suzano mous
gb_est21:AI180114       -  292.00  601.02   9.7e-25   526  ' AI180114 EST222951 Normalized
gb_est39:AW106241       -  291.00  597.41   1.4e-24   577  ' AW106241 um25a09.y1 Suzano mous
gb_est14:AA996890       -  287.50  592.84   2.9e-24   467  ' AA996890 mr38g06.r1 Life Tech
gb_gss1:AC014613        -  274.50  563.42   1.7e-22   559  ' AL005822 F.rubripes GSS sequen
gb_est1:T08254          -  269.00  552.07   5.7e-22   483  ' T08254 ye40g01.r1 Stratagene L
gb_est37:AI976981       -  258.50  530.14   9.2e-21   519  ' AI976981 EST271498 Schistosoma
gb_est11:AA280204       -  246.00  504.59   1.4e-19   407  ' AA280204 zt05c05.r1 NCI_CGAP_Gl
gb_est22:AI006991       -  246.00  507.05   1.4e-19   357  ' AI006991 ua79a04.r1 Soares 2NbM
gb_est30:AI663516       +  243.00  496.19   3.1e-19   557  ' AI663516 uk33g11.y1 Suzano mous
gb_est9:AA096641        +  230.00  469.56   2.4e-17   507  ' AA096641 mo09f08.r1 Life Tech
gb_est1:AA504103        -  226.50  459.44   2.4e-17   275  ' AA504103 nb41g10.s1 NCI_CGAP_Li
gb_est21:AA980800       -  216.50  441.43   8.3e-16   478  ' AA980800 ua45g06.r1 Soares mous
gb_est38:AW052742       -  215.00  435.97   1.6e-15   577  ' AW052742 fi14021B10.y1 614 - roo
gb_est39:AU079281       -  212.00  414.79   6.8e-15   809  ' AU079281 AU079281 Sugano mouse
gb_est32:AI727670       -  211.00  427.74   4.6e-15   561  ' AI727670 BNLGH18551 Six-day Cot
gb_est38:AW044733       -  209.00  422.82   9.7e-15   593  ' AW044733 um13h05.y1 Suzano mous
gb_est16:AI882015       -  206.00  417.65   1.7e-14   526  ' AI882015 ul30c05.y1 Sugano mous
gb_est36:AI931278       -  206.00  417.90   1.6e-14   524  ' AI931278 ul67b04.y1 Suzano moun
gb_est36:AI876404       +  195.00  395.62   2.9e-13   474  ' AI876404 uk74f05.y1 Schiller mo
gb_est12:AV093206       -  194.00  409.10   1.4e-13   212  ' AV093206 RIKEN full-le
gb_est35:AI952403       -  192.50  389.00   4.2e-13   379  ' AI952403 UI-M-BH0-apo-f-05-0-UI
gb_est10:AA129317       -  192.00  387.40   8.0e-13   551  ' AA129317 zn84e05.s1 Stratagene

gb_est44:AW201480       -  187.00  [illegible]  [illegible] ' AW201480 [illegible] ...y1 Gm-c1
gb_gss1:FR0039132       -  186.50  [illegible]  [illegible] ' AL... Fugu rubripes GSS se
gb_est16:299377         -  186.00  [illegible]  [illegible] ' D...77 DKFZphamy1

seq_name: gb_est27:AI437189

seq_documentation_block:
LOCUS       AI437189       499 bp    mRNA            EST       ??-MAR-1999
DEFINITION  fb38g06.y1 Zebrafish Wash[illegible] MPIMG EST [illegible] similar to
            TR:[illegible] DNA sequence.
ACCESSION   AI437189
VERSION     AI437189.1  GI:4288871
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,P., Li,B., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,T., Wylie,T.,
            Underwood,K., Stephen,M., Theising,B., Blair,M., Dennis,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,L., Harvey,N., Schurk,R.,
            Ritter,E., Kuhn,S., Shan,T., Jackson,T., Tejirian,M., M'Timkulu,B.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     On May 18, 1998 this sequence version replaced [illegible].
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. cDNA Sequencing by: Washington University Genome
            Sequencing Center. Clone distribution: I.M.A.G.E. Consortium,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            ResourcenZentrum/Primärbibliothek, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1..499
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish Wash[illegible] MPIMG EST"
                     /sex="mixed"
                     /tissue_type="24 somite embryos, shield
                     stage embryos"
                     /lab_host="[illegible] MRF"
                     /note="Vector: pT[illegible]; Site_1: Not I; Site_2: Sal; 1st
                     strand cDNA was primed with a Not I-[illegible](dT)18 primer
                     ([illegible]ACTTATCGATGTTTTTTTTTTTTTTTTTT); Double-stranded cDNA was [illegible] (BRL),
                     digested with Not I and size[illegible] the Not I and Sal I
                     sites of the pT[illegible] vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab, IMB, Cooke and Max Planck
                     Institute for Molecular Genetik, Berlin). cDNAs for EST
                     analysis were selected following size-[illegible]
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 hrs), adult liver or
                     embryonic shield stage (6 hr) libraries. Fingerprint
                     data were used to computationally [illegible] cDNA and a
                     single cDNA from each [illegible] was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to [illegible] [illegible] [illegible] clones
                     were sequenced [illegible] manual [illegible] assess quality
                     control."
BASE COUNT      163 a    110 c    117 g    109 t
ORIGIN
```

```
alignment_scores:
            Quality:  595.50              Length:    147
              Ratio:  4.347                Gaps:      1
 Percent Similarity:  93.197   Percent Identity: 72.789

alignment_block:
 US-09-174-937-2 x AI437189   ..

 Align seg 1/1  to: AI437189  from: 1  to: 493

     205 ProAspTyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleSerSe 221
         ||:|:|:|| |:|:|:::|:|||:|:|:|:|:|| |||||||||:::|||||
       3 CCAGAIIACCAIIACCIGAIGGAAAAICICIIIICIGIATGTTGTAAGCAG 52

     221 rLeuGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyA 239
         |||||||::: ||||||||:|||||||:::|| :::|||||:|||||||:
      53 TTTCGAGATCTTGGTGGCGAGGATTACCTGATCATCTACATGAATGGAG 102

     239 laThrProArgArgArgMetProGlyIleGlyTrpLeuLysLysCysTyr 254
         ::|||||||||    :::|||||||||||||:::| ||||||||||||||
     103 GAACGCCGCGGAGIAAAATGCCCGGCAITAGCIGGCTCAAGAAGTGCTAC 152

     255 GlnMetIleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleVa 271
         |||||:|||||||||||||||||||||| |||:|||||||||||||||::
     153 CAGATGATCGACAGAAGACTGAGAAAAAACCTAAAATCTCTGATCATCAC 202

     271 lHisProSerTrpPheIleArgThrValLeuAlaIleSerArgProPheI 288
         :|||||:|||||||||||||||||...|| ||||||...|||||||
     203 TCATCCCTCCTGGTTTATTCGGACGGTCATGGTCATTTCCAAACCCTTCA 252

     288 leSerValLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAsp 304
         |:|:|||||:|||||:::|||||||||...||| ||||||||:|||||:::
     253 TCAGCGTGAAGTTCATGAATAAGATCCGTTACGTGCACAGTTTGGAAGAG 302

     305 LeuGluGlnLeuIleProMetGluHisValGlnIleProAspCysValLe 321
         :|||||:::::::::|:|||||:::|||:::... |||||...:|||.:|||
     303 CTGGAAAAGTTCGTTCCCATGGATCATATTCATATCCCAGAATGCATCTT 352

     321 uGlnTyrGluGluGluArgLeuLysAlaArgArgGlu...SerAlaArgP 337
         | |:: | | |:|:|:|||:::|:|| |...|:|          .:.
     353 ACAAIAIGAAGAGGAAGGGITGAACGCTAGAAAAGAGAGAATGGAGCAGG 402

     337 roGlnProGluPheValLeuProArgSerGluGluLysPro 350
           |:|  :||           ::::::::::::::|||||
     403 AACAGAAAGAGCAACAGCAICAAGAACAGCAACAAAAGCCA 443

seq_name: gb_est1:T05645

seq_documentation_block:
LOCUS       T05645        368 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST03534 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDD71, mRNA sequence.
ACCESSION   T05645
VERSION     T05645.1  GI:316795
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699055
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):82344"
                     /db_xref="taxon:9606"
                     /clone="HFBDD71"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: Lambda ZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis, lambda ZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT       88 a    106 c     90 g     80 t      4 others
ORIGIN


alignment_scores:
            Quality:  592.00              Length:    122
              Ratio:  4.975                Gaps:      0
 Percent Similarity:  97.541   Percent Identity: 97.541

alignment_block:
 US-09-174-937-2 x T05645   ..

 Align seg 1/1  to: T05645  from: 1  to: 368

     206 AspTyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLe 222
         ||||||||||||||||||||||||||||||||||||||||||||||||| ||
       2 GACTACCACTACATCATGGAGAACCTCTTCCTGTACGTNATCAGCAGCTT 51

     222 uGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaT 239
         |||||||||||||||||||||||||||||| | |||||||||  |||| |
      52 AGAGCTCCTGGTGGCTGAGGACTACATGATCGTNTACCTGAACGGTGCCA 101

     239 hrProArgArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGln 255
         |||||||||||||||||||||||||| | | ||| |||||| | || | |
     102 CGCCCCGGCGGAGGATGCCTGGAATCGGCTGGCTGAAGAAGTGCTACCAG 151

     256 MetIleAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHi 272
         ||||||||||||   |||||||||||||||||||||||||||||.| |||
     152 ATGATCGACCGGAG.TTGCGGAAAAACCTGAAGTCCTTGATCATCGTCCA 200

     272 sProSerTrpPheIleArgThrValLeuAlaIleSerArgProPheIleS 289
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CCCCTCGTGGTTCATTCGGACTGTGCTGGCCATCTCTCGCCCTTTCATCA 250

     289 erValLysPheIleAsnLysIleGlnTyrValHisSerLeuGluAspLeu 305
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GCGTCAAGTTCATCAACAAGATCCAGTACGTGCACAGCTTGGAAGACCTN 300

     306 GluGlnLeuIleProMetGluHisValGlnIleProAspCysValLeuGl 322
         |||||||||||||||||||||||||||||||||||||||||   |||||||
     301 GAGCAACTCATCCCTATGGAACACGTCCAGATCCCAGACTTCGTCCTGCA 350

     322 nTyrGluGluGluArg 327
         |||||||||||||||
     351 ATACGAGGAGGAAAGA 366

seq_name: gb_est1:T31973

seq_documentation_block:
LOCUS       T31973        423 bp    mRNA            EST       [illegible]95
DEFINITION  EST41767 Human Uterus Homo sapiens cDNA 5' end similar to similar
            to rhoGAP protein, mRNA sequence.
ACCESSION   T31973
VERSION     T31973.1  GI:614071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
```

```
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., , Dillon,P.J., Fannon,M.P., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC14305
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..423
                     /organism="Homo sapiens"
                     /db_xref="ATCC (bacteria) 102311"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Uterus"
                     /note="Organ: uterus"
BASE COUNT      128 a     71 c     90 g    131 t      3 others
ORIGIN


alignment_scores:
            Quality:  496.50              Length:    142
              Ratio:  3.972                 Gaps:      3
 Percent Similarity:  88.028   Percent Identity:  87.696

alignment_block:
 US-09-174-937-2 x T31973   ..

 Align seq 1/1  to: T31973  from: 1  to: 423

     171 IleArgProTyrMetLysValValThrHisGlyGlyTyrTyrGlyGluGl 187
         ||:|| | | |  |||| |:...|||| |  |||||||   ...||
       5 ATTGAACCCIATAAAAAAGTTATCAGCCATGGTGGATATTATNGGGATGG 54

     187 yLeuAsnAlaIleIleValPheAlaAlaCysPheLeuProAspSerSerL 204
         |||:|||||||:::||||||||...|| |  ...|||...|||||||
      55 ATTAAATGCCATTGTTGTGTTTGCTGTCTGTT.CATGCCTGAAAGTAGTC 104

     204 euProAspTyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleSer 220
         || ||||||:|||| ||||||||:|||||.|  |   |||||||||:::
     105 AGCCTAACTATAGATACCTGATGGACAATCTT.TTAAATATGTTATTGGC 154

     221 SerLeuGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsnGl 237
         ::: ' ' ' : |||| | |||:::|| |  |  |||||||| ||||||
     155 ACTTTGGAGCTATTAGTAGCAGAAAACTACATAGATAGTTATTTAAAATGG 204

     237 yAlaThrProArgArgArgMetProGlyIleGlyTrpLeuLysLysCysT 254
         |'|||||  |||||||:::|||||||:::::|| |||||||:::|||||||
     205 TGCAACAACTTGAAGAAAAATGCCCAGTCTGGGATGGCTCAGAAATGTT 254

     254 yrGlnMetIleAspArgArgLeuArg.LysAspLeuLysSerLeuIleIl 270
         ' :::' ' ||||| | ' ' ||| ||'     ||||||| |||'|||
```

```
     255 ATCAGCAAAATTGATAGAAGGTTACGGAAAAATCTAAATCTTTAATTAT 304

     270 eValHisProSerTrpPheIleArgThrValLy AlaIleSerArg.Pro 29
         .:' |:|||||||'  |'| ::: :  :::::
     305 TTTACATCCTTTTGGgttATCAGAAGAATTCTTGCTGTTANAA'''''' 354

     287 PheIleSerValLysPheIleAsnLysIleGlnTyrVa HisArgA 301
                 :::'' '                    ||:
     355 TTTATTAGCTCGGAATTTCAGCCAAAAAATTTGGTAC...GGTTTTATTTN 4 1

     303 uAspLeuGluGlnLeuIlePro 31
         :::  ::::::::::::'''
     402 GGAGACTAGCAGACTTGTCCCC 423

seq_name: gb_est1:T08680

seq_documentation_block:
LOCUS       T08680         437 bp    mRNA             EST         -APR-1995
DEFINITION  EST06572 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBI25
            5' end, mRNA sequence.
ACCESSION   T08680
VERSION     T08680.1  GI:389708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..437
                     /organism="Homo sapiens"
                     /db_xref="ATCC (bacteria):102311"
                     /db_xref="taxon:9606"
                     /clone="HIBBI25"
                     /clone_lib="Infant Brain, Bento Soares"
BASE COUNT      117 a    127 c    117 g     84 t      2 others
ORIGIN


alignment_scores:
            Quality:  449.00              Length:     8
              Ratio:  5.045                 Gaps:
 Percent Similarity: 100.000   Percent Identity: 1.  .

alignment_block:
 US-09-174-937-2 x T08680   ..

 Align seq 1/1  to: T08680  from: 1  to: 437

     283 IleSerArgProPheIleSerValLysPheIleAsnLysIleGlnTyrVa 299
           | | | |  |  |  |
       3 ATCTCTCCCCCTTTCATCAGCCTCAATTCATCAATAAGATACAGTATGTA 52

     299 lHisSerLeuGluAspLeuGlnLeuIleProMetGlnHisVal 316
         ' ' ' ||||||||||| ' '
      53 GCACAGCTTCGAAGACCTGCAGCAATTCATCCCATTGCAACATGTCAGA 102

     316 leProAspCysValLeuGlnTyrGlnGluArgLeuLysAlaArgArg 332
         |||| | |||||||||| ||||  ' ' ' ' ' ' ' ' ' ' ' '
     103 TCCCAGACTGCGTCCTGCAATACCAAGAGAGGTTAAAAGCTCGAAGA 150
```

```
     333 GluSerAlaArgProGlnProGluPheValLeuProArgSerGluGluLy 349
         |.|||.|'||||||||||||||||||||||||.||  |||||||||.|||||||
     153 GAGAGCGCGAGGCCCCAGCCGGAGTTTGTGCTGTCCAGGTCTGAAGAGAA 202

     349 sProGluValAlaProValGluAsnArgSerAlaLeuValSerGluAspG 366
         ' | | | . | ' | | | |'|'|||| ||  '||||||| |||||||
     203 GCCAGAGGTGGCACCAGTGGAAAACAGGTCTGTCTGGTCTCAGAAGATC 252

     366 lnGluThrSerMetSer 371
         ' '|'''''|'|||'
     253 AGGAAACAAGCATGTCC 269

seq_name: gb_est10:AA137873

seq_documentation_block:
LOCUS       AA137873      449 bp    mRNA            EST       09-FEB-1997
DEFINITION  mq75d03.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:584549 5' similar to TR:G558844 G558844 NIP2. , mRNA
            sequence.
ACCESSION   AA137873
VERSION     AA137873.1  GI:1700244
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 10, 1996 this sequence version replaced gi:1394228.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1900
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:359197
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:584549"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SURE (tetracycline resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally; Primer:
                     Oligo dT; From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma; Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; 5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      147 a     76 c    124 g    102 t
ORIGIN


alignment_scores:
            Quality:  426.50            Length:    121
              Ratio:  4.101               Gaps:      2
 Percent Similarity:  85.950   Percent Identity: 63.636

alignment_block:
 US-09-174-937-2 x AA137873   ..
```

```
 Align seq 1/1  to: AA137873  from: 1  to: 449

     142 ThrThrGluAspGlySerAlaAlaAsnGlyArgLeuTrpArgThrVal11 184
         |||   |||   |||          :::|||||   |||
      96 ACAGAAGAAAAGGGC.........GATGGACGCCGCTGGCCAATGTTCAG 146

     158 eIleGlyGluGlnGluHisArgIleAspLeuHisMetIleArgProTyrM 175
          ||||||||||||:..||||||.::|||||:      |||||: : ||
     137 GATTGGAGAGCAGGACCACAGGGTGGACATGAAGGCCATCGAGCCCTACA 186

     175 etLysValValThrHisGlyGlyTyrTyrGlyGluGlyLeuAsnAlaIle 191
         ||||||.....:|||||||||||'| |||||::|||||'||||||||||
     187 AAAAAGTTATTAGCCATGGAGGATATTATGGGGACGGCTTAAATGCCATT 236

     192 IleVal PheAlaAlaCysPheLeuPro AspSerSerLeuProAspTyrM 208
         ...||| ||||||...||||||:::| |:::||||::   |||:::|||:
     237 GTTGTTATTTGTTGTCTGTTTCATGCCTGAGAGTTGCTAGCCTTAACTATA 286

     208 isTyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluGlu 224
         ..||| .. ||| ..|||||||||     ||||||:::::|'|:|||
     287 GATATTTGATGGACAATCTCTTTAAATATGTTATTGTGCACTTTAGAGCTG 336

     225 LeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProAr 241
         ||||||||||||..|||||||||':|||||||||||||||'||||'   ||
     337 TTGGTAGCAGAGAACTACATGATCATTTACTTAAATGGTGCCACAACTCCG 386

     241 gArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleA 257
         ||||:::|||||||'''''|||||||||:::::|||||||||:::|'||.
     387 AAGAAAAATGCCCAGTTTGGGATGGCTACGAGATGCTACCAGCAAATTG 436

     258 spArgArgLeu 261
         ||||||||||
     437 ATAGACGATTA 447

seq_name: gb_est1:Z32816

seq_documentation_block:
LOCUS       Z32816        312 bp    mRNA            EST       08-MAY-1994
DEFINITION  HHEA02P Atrium cDNA library Human heart Homo sapiens mRNA clone
            HEA02P, mRNA sequence.
ACCESSION   Z32816
VERSION     Z32816.1  GI:483499
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Rougeot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE, molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III (Sci. Vie) 319 (2), 263-272 (1996)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276158.
            Contact: Genzentrum Muenchen
            Laboratorium fuer molekulare Biologie
            Am Klopferspitz 18a,8033 Martinsried,Germany
            Email: obermaier@vms.biochem.mpg.de
            single read.
FEATURES             Location/Qualifiers
     source          1. .312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEA02P"
                     /clone_lib="Atrium cDNA library Human heart"
BASE COUNT       97 a     55 c     59 g    101 t
ORIGIN
```

```
alignment_scores:
            Quality:  422.00              Length:    102
              Ratio:  4.351                Gaps:      0
 Percent Similarity:  95.098   Percent Identity: 75.490

alignment_block:
 US-09-174-937-2 x 232816   ..

 Align seq 1/1  to: 232816  from: 1  to: 312

     191 IleIleValPheAlaAlaCysPheLeuProAspSerSerLeuProAspTy 207
         ||||:|||||||||||:::|||||||:::|||:::: |||||   |||:::|| 
       1 ATTTTTGTGTTTGCTGTCTGTTTCATGCCIGAAAGTAGTCAGCCTAACTA 50

     207 rHisTyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluL 224
         |:::|||:::|||:::|||||||||   ||:: ||||:::::.|||||| 
      51 TAGATACCTGATGGACAATCTTTTTAAATATGTTATTGGCACTTGGAGC 100

     224 euLeuValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrPro 240
         ||||||||||||||:::|||||||||||||||:|||||||||||||||| 
     101 TACTAGTAGCAGAAAACTACATGATAGTTTATTTAAATGGTGCAACAACT 150

     241 ArgArgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIl 257
         ||||||:::|||||:::::|||||||||:::   ||||||||:::||| 
     151 CGAAGAAAAATGCCCAGTCTGGGATGGCTCAGGIAATGTTATCAGCAAAT 200

     257 cAspArgArgLeuArgLysAsnLeuLysSerLeuIleIleValHisProS 274
         ||||||||||||||||:::|||||||||||.|:::|||||||||||||| 
     201 TGATAGAAGGTTACGGGAAAATCTAAAATCCCTAATCATTGTACATCCTT 250

     274 erTrpPheIleArgThrValLeuAlaIleSerArgProPheIleSerVal 290
         |||||||||||||||||:::|||||:::::|||||||||||||| 
     251 CTTGGTTTATCAGAACACTTCTGGCTGTTACAAGACCATTTATTAGCTCG 300

     291 LysPhe 292
         |||||| 
     301 AAATTC 306

seq_name: gb_est14:AA450831

seq_documentation_block:
LOCUS       AA450831      368 bp    mRNA            EST       04-JUN-1997
DEFINITION  vg55e03.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865276 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA450831
VERSION     AA450831.1  GI:2164501
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394794.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509364
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 332.

FEATURES             Location/Qualifiers
     source          1..368
                     /organism="Mus musculus"
                     /strain="C57BL6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:865276"
                     /clone_lib="Beddington mouse embryonic region"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT. Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings, excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                     Referenced in Development 121: 2479-2489 (1995)"
BASE COUNT      110 a      66 c      98 g      94 t
ORIGIN


alignment_scores:
            Quality:  419.00              Length:    1 2
              Ratio:  4.365                Gaps:      2
 Percent Similarity:  94.118   Percent Identity: 68.627

alignment_block:
 US-09-174-937-2 x AA450831   ..

 Align seq 1/1  to: AA450831  from: 1  to: 368

     159 IleGlyGluGlnGluHisArgIleAspLeuHisMetIleArgPr.TyrMe 175
         ||||:|||||||:::|||||:::| :::    ' :::    ::  
      63 ATTGGAGAGCAGGACCACAGGGTGACATGAAGCGATCGAGCCTTACAA 112

     175 tLysValValThrHisGlyGlyTyrTyrGlyG:uGlyLeuAsnA.alleI 192
         |||||:::::|||||||||.| ' :::  ::: | '    :::  :  
     113 AAAAGTTATTAGCCATGGAGGAIA:IATGGGGAIGGCTTAAATGTCATTA 162

     192 leValPheAlaAlaCysPheLeuProAspSerSerLeuProAspTyrHis 2.9
         :|||||||||||||.|| |||:: ' :::  :::    :::  :  
     163 TTGTATTTGCTGTCTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAA 212

     209 TyrIleMetGluAsnLeuPheLeuTyrValIleSerSerLeuGluIleIle 225
         |||:::|||:::|||||||||    ' ' ' ::::::  
     213 TATTTGATGGACAATCTCTTTAAATATGTTATTGGCACTTTAGAGCTCTT 262

     225 uValAlaGluAspTyrMetIleValTyrLeuAsnGlyAlaThrProArgA 242
         |||| ||||:::|||| ||  '|:::  ' ' |  '
     263 GGTAGCAGAGAACTACATGATCATTTATTTAAATGGTGCGACAACTCGAA 312

     242 rgArgMetProGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAsp 258
         |:::|'|||:....:'|'||| ' :::::  ' ' ' :::  :
     313 GAAAAATGCCCAGTTTGGGATGGCTCAGGATATCCTGCAGCAAAATTGAT 362

     259 ArgArg 260
         ||:|'|
     363 AGACGA 368

seq_name: gb_est9:AA067188

seq_documentation_block:
LOCUS       AA067188      452 bp    mRNA            EST       16-FEB-1997
DEFINITION  mm38b12.r1 Stratagene mouse melanoma (#937212) Mus musculus cDNA
            clone IMAGE:523775 5' similar to TR:G558844 G558844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA067188
VERSION     AA067188.1  GI:1564954
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,J., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393488.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:317623
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:523775"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      145 a     90 c    108 g    109 t
ORIGIN


alignment_scores:
           Quality:  402.00            Length:    134
             Ratio:  3.655               Gaps:      3
 Percent Similarity:  82.090   Percent Identity: 57.463

alignment_block:
 US-09-174-937-2 x AA067188   ..

 Align seq 1/1  to: AA067188  from: 1  to: 452

     233 ValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyIleGlyTr 249
         :::|||'||||||'|.||||   ||||| ::: |'|||.....|||||
       1 ATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGATG 50

     249 pLeuLysLysCysTyrGluMetIleAspArgArgLeuArgLysAsnLeuL 266
         '''|::::::'|||||'||'|::::|||||| ||  |||||||||'|||||
      51 GCTCAGGAGATGCTACCAGCAAATTGATAGACGATTAAGGAAAAACCTTA 100

     266 ysSerLeuIleIleValHisProSerTrpPheIleArgThrValLeuAla 282
         |||||'|'|||||||||||||||||||||||.||  ||||||||:::|||||
     101 AGTCTCTAAICATTGTGCACCCCTCTTGGTTTATCCGAACACTTCTGGCT 150

     283 IleSerArgProPheIleSerValLysPheIleAsnLysIleGlnTyrVa 299
         :::::::'|||||||||||'   ||||   :::|||||||:::|||||
     151 GTTACACGACCATTTATCAGTTCAAAGTTCAGTCAGAAGATCAGATACGT 200

     299 HisSerLeuGluAspLeuGluGlnLeuIlePr MetGluHisValGlnI 316
         |  :::|||   :::|||   :::|||||::|'|||||||:::|||   |
     201 CTTTAACTTGGCCGAGCTGGCAGAGCTTGTTCCATGGAGTATGTCGGCA 250

     316 lePr AspCysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArg 332
         |||| ::::|:::   ||||||||||||.|||::::::|||   |||:::
```

```
     251 TACCAGAGTTGCATCAAACAGTATCAACAACAAAAGTTTAAAAAAAAAA 300

     333 Glu . .                .SerAlaArgProGlnProGluPheValIle 343
         ...                    .:::::::   :::|||
     301 AAAAGAGTTGATCAAGAGCTTAATGGAAAACAAGAACCA..........  344

     343 uProArgSerGluGlu LysProGluValAlaProValGluAsnArgSer 359
          |||:::|||||||::: |||   ::: ||   |||   ::::::|||:::
     345 CCAAAAAGTGAGCAGTAAGTTCCCTGTCCAGCCGGACAAGAGAGAGCA 393

seq_name: gb_est9:AA068579

seq_documentation_block.
LOCUS       AA068579      443 bp    mRNA            EST       26-FEB-1997
DEFINITION  mm45e06.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
            clone IMAGE:524482 5' similar to TR:G559944 G559944 MIP2 ., mRNA
            sequence.
ACCESSION   AA068579
VERSION     AA068579.1  GI:1575939
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,J., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802463.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:318330
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1..443
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:524482"
                     /clone_lib="Stratagene mouse melanoma (#937312)"
                     /tissue_type="melanoma"
                     /dev_stage="M2 cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. From M2 cells, a highly metastatic derivative of
                     the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      143 a     89 c    106 g    105 t
ORIGIN


alignment_scores:
           Quality:  402.00            Length:    134
             Ratio:  3.655               Gaps:      3
 Percent Similarity:  82.090   Percent Identity: 57.463

alignment_block:
 US-09-174-937-2 x AA068579   ..

 Align seq 1/1  to: AA068579  from: 1  to: 443
```

```
      233 ValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyIleGlyTr 249
          :::|||||||||||||||||    ||||||:::|||||:::::|||||
        1 ATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCCCAGTTTGGGATG 50

      249 pIeuLysLysCysTyrGlnMetIleAspArgA:gLeuArgLysAsnLeuL 266
          |||:::::: ||||||||||:::|||||||| | :|||||||||||||||
       51 GCTCAGGAGATGCIACCAGCAAATTGATAGACCATTAAGGAAAAACCTTA 100

      266 ysSerLeuIleIleValHisProSerTrpPheIleArgIhrValLeuAla 282
          |||||||||||||||||||||||||||||||:|' ||||||||::||||||
      101 AGTCTCTAATCATTGTGCACCCCTCTTGGTTTATCCGAACACTTCTGGCT 150

      283 IleSerArgProPheIleSerValLysPheIleAsnLysIleGlnTyrVa 299
          :::::|||||||||||||||||    ||||||    ..:|||||||:::|||||
      151 GTTACACGACCATTTATCAGTTCAAAGTTCAGTCAGAAGATCAGATACGT 200

      299 lHisSerLeuGluAspLeuGluGlnLeuIleProMetGluHisValGlnI 316
          |    ::|||||   ...|||   :::||:::::. ||||||:::|||    |
      201 CTTTAACTTGGCCGAGCTGGCAGAGCTTGTCCCATGGAGTATGTCGGCA 250

      316 leProAspCysValLeuGlnTyrGluGluGluArgLeuLysAlaArgArg 332
          ||||:::|||:::   ||||||||||||| |:::::|||   ||||::
      251 TACCAGAGTGCATCAAACAGTATGAAGAAGAAAAGTTTAAAAAGAGACAA 300

      333 Glu..  .  ..........SerAlaArgProGlnProGluPheValLe 343
          :::      :.........  :::|||
      301 AAAAGAGTTGATCAAGAGCTTAATGGAAAACAAGAACCA.......... 339

      343 uProArgSerGluGlu.LysProGluValAlaProValGluAsnArgSer 359
          |||:::|||||||:: |||   :::|||   :||   ::::::|||:::
      340 .CCAAAAAGTGAGCAGTAAGTTCCGTGTCCAGCCCGGACAAGAGAGAGCA 388

seq_name: gb_est21:AA914577

seq_documentation_block:
LOCUS       AA914577      516 bp    mRNA            EST       14-APR-1998
DEFINITION  vs02e07.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1314564 5' similar to TR:Q12982 Q12982 NIP2. ;, mRNA
            sequence.
ACCESSION   AA914577
VERSION     AA914577.1  GI:3053969
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407399.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email. mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:684860
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop. Sss.
FEATURES             Location/Qualifiers
     source          1..516
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1314564"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer (5'
                     IGTTACCAATCTCAAGTGGGAGCGTCCCCCAATGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'); double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia); digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Benito Soares and M.Fatima
                     Bonaldo."
BASE COUNT      148 a    110 c    118 g    141 t
ORIGIN


alignment_scores:
            Quality:  392.00              Length:    1:2
              Ratio:  3.630                 Gaps:      3
 Percent Similarity:  81.818   Percent Identity: 57.576

alignment_block:
 US-09-174-937-2 x AA914577   ..

 Align seg 1/1  to: AA914577  from: 1  to: 516

      229 AspTyrMetIleValTyrLeuAsnGlyAlaThrProArgArgArgMetPr 245
          :::|||||||||:::|'|||||  |  '  '  |  ::  '  '  '  '
        2 AACTACATGATCATTTACTTAAATGGTGCGACAACTCGAAGAAAAATGCC 51

      245 oGlyIleGlyTrpLeuLysLysCysTyrGlnMetIleAspArgArgLeuA 262
          |:::::::::::::::::::|'  '  '  ::  '  '  '  '  '  '
       52 CAGTTTGGGATGGCTCAGGAGATGCTACCAGCAAATTGATAGACCATTAA 1.1

      262 rgLysAsnLeuLysSerLeIleIleValHisProSerTrpPheIleArg 278
          |||||||||||||||| ||||' '  '  '  '  '  '  '  '  '  '
      102 GGAAAAACCTIAAGTCTCTAATCATTGTGCACCCCTCTTGGTTTATCCGA 151

      279 ThrValLeuAlaIleSerArgProPheIleSerValLysPheIleAsnLy 295
          |'|:::|||||:::::::|||:  '  '  '  '  '  '  '  :::
      152 ACACTTCTGGCTGTTACACGACCA:::A'CAGTTCAAAGTTCAGTCAGAA 2.1

      295 sIleGlnTyrValHisSerLeuGluAspLeuGluGlnLeuIleProMetG 312
          |||:::|||||||   :::  ||   :::  '   :::  :::  '  '
      202 GATCAGATACGTCTTTAACTTGGCTGAGCTGGCAGAGCTTGTCCCATGGA 251

      312 luHisValGlnIleProAspCysValLeuGlnTyrGluGluGluArgLeu 328
          |':::::'   '|||||:::|':::  '  '  '  '  ::::::  :::
      252 AGTATGTCGGCATACCAGAGTGCATCAAACAAGTTGATCAAGAG...GTT 299

      329 LysAlaArgArgGluSerAlaArgProGlnProGluPheValIleArgAr 345
          ::::::::::::::|||  '  '  '  '  '  '  '  '  '  '  ::
      299 AATGGAAAACAAGAACCA.....................CCAAA 321

      345 gSerGluGlu.LysProGluValAlaProValGluAsnArgSer 359
          :|'|||||:::.|||   :::|||'  '  '  '  ::::::  :::  '
      322 AAGTGAGCAGTAAGTTCCGTGTCCAGCCCGGACAAGAGAGAGCA 365

seq_name: gb_est9:AA113114

seq_documentation_block:
LOCUS       AA113114      382 bp    mRNA            EST       01-AUG-1997
DEFINITION  zn69h02.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
            IMAGE:563379 5' similar to TR:O55844 O55844 NIP2. ;, mRNA
            sequence.
ACCESSION   AA113114
VERSION     AA113114.1  GI:1664684
KEYWORDS    EST.
```

```
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On May 5, 1995 this sequence version replaced gi:797955
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2034   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1..382
                     /organism="Homo sapiens"
                     /db_xref="GDB:4583190"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:563379"
                     /clone_lib="Stratagene HeLa cell s3 937216"
                     /sex="female"
                     /dev_stage="HeLa S3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI. Cloned unidirectionally. Primer: Oligo dT. HeLa S3
                     epithelioid carcinoma cells grown to semi-confluency
                     without induction. Average insert size, 1.5 kb, Uni-ZAP XR
                     Vector. -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3'
                     adaptor sequence: 5' CTCGATTTTTTTTTTTTTTTT 3'"
BASE COUNT      131 a     59 c     93 g     99 t
ORIGIN


alignment_scores:
            Quality:  390.00            Length:    125
              Ratio:  3.679              Gaps:      3
 Percent Similarity:  84.800   Percent Identity: 59.200

alignment_block:
 US-09-174-937-2 x AA113114   ..

 Align seq 1/1  to: AA113114  from: 1  to: 382

     134 SerAlaAspLeuPheGlyAspGlyThrThrGlnAspGlySerAlaAla.. 149
         ::::::::::::::::     ||||:.      ...    . :|||||
      11 ACTACTGAAGTAATTAGGAAAGGGTCAATTACTGAATACACAGCAGCAGA 60

     150 ..........AsnGlyArgLeuIleArgThrValIleIleGlyGlnGln 162
                  :::|||||||   ||||||:::||.    ||||||||||||
      61 GGAAAAAGAAGATGGACGACGTGGCGTAGTGTTCAGGATTGGAGAACAG 110

     163 GluHisArgIleAspLeuHisMetIleArgProTyrMetLysValValTh 179
         ::|||||||||:::|||...     |||...||.||   ||    ||||||.....
     111 GACCACAGGGTTGATATGAAGGCAATTGAACCTATAAAAAAGTTATCAG 160

     179 rHisGlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleValPheAlaA 196
         :|||||*|||||||||||||:::|||||| |   |||:::|||||||||:
     161 CCATGGGGGATATTATGGGGATGGATTAAAIGCAATTGTTGTGTTTGCTG 210

     196 laCysPheLeuProAspSerSerLeuProAspTyrHisTyrIleMetGlu 212
         ..||||||...|||...||||||   |||:::|||:::|||:::|:|:::
     211 TCTGTTTCATGCCTGAAAGTAGTCAGCCTAACTATAGATACCTGATGGAC 260

     213 AsnLeuPheLeuTyrValIleSerSerLeuGluLeuLeuValAlaGluAs 229
         |||||||||   |||||||||:::::|||:|||||||:|||:|||:::
     261 AATCTTTTTAAATATGTTATTGGCACTTTGGAGCTATTAGTAGCAGAAAA 310

     229 pTyrMetIleValTyrLeuAspGlyAlaThrProArgArgArgMetProG 246
         :|||||||||||||||||||||||||||||:|||||   ||||:::::::::
     311 CTACATGATAGTTTATTTAAATGGTGCAACAACTCGAAGAAAAATGCCTA 360

     246 lyIleGly.TrpLeuLysLys 252
         :.|||.|||||  :|||
     361 GTCTGGGAATGGCTCAGGAAA 381

seq_name: gb_est29:AI620433

seq_documentation_block:
LOCUS       AI620433      486 bp    mRNA            EST       21-APR-1999
DEFINITION  tu47b05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2254161 3'
            similar to TR:Q12982 Q12982 NIP2., mRNA sequence.
ACCESSION   AI620433
VERSION     AI620433.1  GI:4629559
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2449857.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2254161"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr28 was prepared, and ss
                     circles were rescued with VCS-M13 helper phage. This
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clone IDs
                     985608-986759, 1101102-1101959, and 1217894-1220415).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      119 a    117 c    120 g    130 t
ORIGIN


alignment_scores:
            Quality:  368.00            Length:    141
              Ratio:  3.228              Gaps:      0
```

Percent Similarity: 80.851 Percent Identity: 46.099

alignment_block:
US-09-174-937-2 x AI620433/rev ..

Align seq 1/1 *: reverse of AI620433 from: 1 to: 486

```
182 GlyTyrTyrGlyGluGlyLeuAsnAlaIleIleValPheAlaAlaCysPh 198
    ||||||'     :::|||||||:|||    |:::::::|||:|||:
436 TTTCATTTTTGAATTTTTTTAATTGTGCTAATTTTTTGGCTTCCTGTTA 437

198 eLeuProAspSerSerLeuProAspTyrHisTyrIleMetGluAsnLeuP 215
    :|||||||          :::|||          :::|||||||:::   |
436 TCTACCAAAAACCAAGAATCCCCACCACCCTTTGTGTCATGGAACACTGGT 387

215 heLeuTyrValIleSerSerLeuGluLeuLeuValAlaGluAspTyrMet 231
    ||   |||:::::::.:.:::::.:::::::.:|:|:|:...:|||:::
386 TTAGGTATATGGTGGGAACTTTGAAGCTGCTAGCTGAAAATTACCTG 337

232 IleValTyrLeuAsnGlyAlaThrProArgArgArgMetProGlyIleGl 248
    ':::::::::::'  '||   |'|   |    :  :::|||:    ::
336 CTTGTTCATTTGAGTGGAGGCACAAGCAGGGCCCAAGTTCCACCTTAAAG 287

249 yTrpLeuLysLysCysTyrGlnMetIleAspArgArgLeuArgLysAsnL 265
    :|||:::::::::::|||||:::   :::|||||||||||||||||||||
286 CTGGATACGTCAGTGTTACCGTACCCTGGATCGGCGGCTACGGAAAAACC 237

265 euLysSerLeuIleIleValHisProSerTrpPheIleArgThrValLeu 281
    '||.....|||  :::::|||||:   :::: |:::.:::::.:   |||
236 TGCGAGCCCTGGTGGTTGTCCATGCTACATGGTATGTGAAAGCATTTCTG 187

282 AlaIleSerArgProPheIleSerValLysPheIleAsnLysIleGlnTy 298
    '|:::   :|||||||||||||||   ||||||   :::|||||:::::
186 GCACTGCTTCGGCCCTTCATCAGTTCCAAATTCACACGAAAAATCCGTTT 137

298 rValHisSerLeuGluAspLeuGluGlnLeuIleProMetGluHisValG 315
    ::::   |'|||||   :::|||   |||||||:.    :::: :::|||:
136 TCTGGACAGCCTGGGGGAGCTGGCCCAACTCATATCCCTGGATCAAGTCC 87

315 lnIleProAspCysValLeuGln 322
    ::|||||||:::::::||   |||
 86 ACATCCCTGAAGCTGTCAGACAG 64
```

seq_name: gb_est1:T31983

seq_documentation_block:
LOCUS T31983 281 bp mRNA EST 06-SEP-1995
DEFINITION EST41923 Human Uterus Homo sapiens cDNA 5' end similar to None, mRNA sequence.
ACCESSION T31983
VERSION T31983.1 GI:614081
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 281)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,I.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., , Dillon,P.J., Fannon,M.P., Rosen,C.A.,
Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million [illegible] of cDNA Sequence
JOURNAL Nature 377, 3-174 (1995)
MEDLINE 96026280
COMMENT Other_ESTs: THC14305
Contact: Venter, JC
The Institute for Genomic Research
[illegible] Clopper Rd., Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability and additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
FEATURES Location/qualifiers
source 1..281
/organism="Homo sapiens"
/db_xref="[illegible]"
/db_xref="taxon:9606"
/clone_lib="Human Uterus"
/note="Organ: uterus"
BASE COUNT 89 a 41 c 63 g 86 t 2 others
ORIGIN

alignment_scores:
Quality: 357.00 Length: 1
Ratio: 4.250 Gaps: 1
Percent Similarity: 92.308 Percent Identity: [illegible]

alignment_block:
US-09-174-937-2 x T31983 ..

Align seq 1/1 to: T31983 from: 1 to: 281

```
171 IleArgProTyrMetLysValValThrHisGlyGlyTyrTyrGlyGlyGl 187
    |'|:::|'|||:   :|||||:::::: | |  ' |  '  '  :::
  5 ATTGAACCCTATAAAAAAGTTATCAGCCATGGGGGATATTATNGGATGG 54

187 yLeuAsnAlaIleIleValPheAlaAlaCysProLeuProAspSerSerL 204
    |||||||||||||||:::'|||||| ::::|  '  :::  :::  ' |
 55 ATTAAATGCCATTGTTGTGTTTGCTTGTCTGTTTCATGCCTGAAAGTAGTC 104

204 euProAspTyrHisTyrIleMetGluAsnLeuPheLeuTyrValIleSer 22
    ||:::|||||:::|||:: |||:::  ' |  '  '  :::  ::::
105 AGCCTAACTATAGATACCTGATGGAAAATCTTTTTAAATATGTTATTGGC 154

221 SerLeuGluLeuLeuValAlaGluAspTyrMetIleValTyrLeuAsnGl 237
    :::|||||||||||||||| ||| :::  '  '  '  '  '  '  '
155 ACTTTGGAGCTATTAGTAGCAGAAAACTACATGATAGTTTATTTAAAATGG 204

237 yAlaThrProArgArg.ArgMetProGlyIleGlyTrpLeuLysLysCys 253
    |||.|||   '|||||| ::|||  ' |:::::  '  '  '  :::
205 TGCAACAACTCGAAGAAAAAATGCCCAGTTGGGATGGCTCAAGNAATGT 254

254 TyrGlnMetIleAspArgArg 269
    |'|'|:::|:|'|||||||||
255 TATCAGCAAATTGATAGAAGG 275
```

seq_name: gb_est10:AA129380

seq_documentation_block:
LOCUS AA129380 394 bp mRNA EST [illegible]-AUG-1997
DEFINITION zn84e05.r1 Stratagene lung carcinoma 937210 Homo sapiens cDNA clone IMAGE:564896 5' similar [illegible] mRNA sequence.
ACCESSION AA129380
VERSION AA129380.1 GI:1689145
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397635.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63110
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 825  Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 163.
FEATURES             Location/Qualifiers
     source          1..394
                     /organism="Homo sapiens"
                     /db_xref="GDB:459960"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:564896"
                     /clone_lib="Stratagene lung carcinoma 937218"
                     /tissue_type="lung carcinoma"
                     /cell_line="NCI-H69"
                     /dev_stage="cell line NCI-H69"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Small cell carcinoma cell line NCI-H69.  Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      114 a     75 c     87 g    113 t      5 others
ORIGIN


alignment_scores:
            Quality:  342.00            Length:    116
              Ratio:  3.526               Gaps:      6
 Percent Similarity:  82.621   Percent Identity: 64.655

alignment_block:
 US-09-174-937-2 x AA129380  ..

 Align Seq 1/1  to  AA129380  from  1  to  394

     179 ThrHisGlyGlyTyrTyrGlyGluGlyLeuAsnAlaIleIleValPheAl 195
         :::|||||||||||||||||||:::|||| || ||||||:::|||||||
           3 AGCCATGGGGGATATTATGGGGATGGATTAAATCCCATTGTTGTGTTTGC 52

     195 aAlaCysPheLeuProAspSerSerLeuProAspTyrHisTyrIleMetG 212
         |:::|||||:::|||:::||||| | |||:::|||..:|||..:|||.
      53 TGTCTGTTTCATGCCTGAAAGTAGTCAGCTTACTATAGATACCTGATGG 102

     212 luAsnLeuPheLeuTyrValIleSerSerLeuLeuLeuValAlaGlu 228
         :::|||||||   |||||||||:::::::::|||||||||||||||||
     103 ACAATCTTTTTAAATATGTTATTGGCACTTTGGAGCTACTAGTAGCAGAA 152

     229 AspTyrMetIleValTyrLeuAsnGlyAlaThrPro.ArgArgArgMetP 245
         :::|||||||||||||||||||||||||||||||:    |||||||:::|||
     153 AACTACATGATAGTTTATTTAAATGGTGCAACAACTCCGAAGAAAAATGC 202

     245 roGlyIleGlyTrpLeuLysLysCysTyrGlnMet.IleAspArg.ArgL 261
         ||:::..:|||||||||...|||||||||||::: |||||||||| ||||
     203 CCAGTCTGGGATGGCTCAGGAAATGTTATCAGCAAAATTGATAGAAAGGT 252

     261 eu.ArgLysAsnLeuLysSerLeuIle.IleValHisProSerTrpPheI 277
         || |||||||||||||||:::  ||| |||  ||||||'||
     253 TACCGGAAAAATCTAAAATNCCCAATCCATTGGACATCCNICC..... 295

     277 leArgThrValLeuAlaIleSerArgProPheIleSer 289
               |||:::  ':||||:::::::
     296 ..........TTGGGTTATCCAGGACCCTTCCTGGCT 322

seq_name: gb_est32:AI722140

seq_documentation_block:
LOCUS       AI722140      604 bp    mRNA            EST       [17-JUN-1999
DEFINITION  fd19g03 y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            2352052 5' similar to TR:[illegible] [illegible] PIR:A37367 [illegible] cDNA sequence.
ACCESSION   AI722140
VERSION     AI722140.1  GI:5040469
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Sugano,S., Kawakami,K., Johnson,S.L., Li,P., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Yost,S., Kucaba,T.,
            Martin,J., Pape,D., Stephens,M., Underwood,K., Theising,B.,
            Pitter,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:234837.
            Other_ESTs: fd19g03.x1
            Contact: S.L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="2352052"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage resistant)"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTCG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACCATGTG; 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to remove fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami.  Custom primers for
                     sequencing: 5' end primer GTTGTGGCCTAAAAGGTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA.  "
BASE COUNT      194 a    145 c    153 g    121 t      1 others
ORIGIN


alignment_scores:
```

```
           Quality:  326.50              Length.    190
             Ratio:  2.531                 Gaps:      6
 Percent Similarity:  67.895   Percent Identity: 38.947

alignment_block:
 US-09-174-937-2 x AI722140   ..

 Align seq 1/1  to: AI722140  from: 1  to: 604

      14 AspValLysGluGluTrpGlnAspGluAspLeuProArgProLeuProGl 30
         |||:::|||   :::   ::::::  :::'|||:::|||      ::
      72 GATCTAAAACCCAGAGAAGGGAACAGAGGTGTGCCTCAACCGTGGTGTTC 121

      30 uGluThrGly...          ... ..........ValGluLeuLeuG 38
         :  :::|||                            ::::::
     122 CCCTTCTGGCAATGTGGGGATGGACCAGACTTATGAAAGCAAGATGAGTT 171

      38 lySerProValGluAspThrSerSerProProAsnThrLeuAsnPhe... 53
         :::  :::|||:::      :::|||||:|::::::|||   :::
     172 TTGAAGGGGCAGAGAGTAGACCAGCTCCACCCACCTCACTGCCTCTACAG 221

      54 ......AsnGlyAlaHisArgLysArgLysThrLeuValAlaProGluIl 68
            :::|||:::      :::|||||   |||   ||||||:::||
     222 GCCCAAGGGGGTGTGTCTTCTCAGAGGAAAAAACTGTCCGCTCCCCGTAT 271

      68 eAsnIleSerLeuAspGlnSerGluGlySerLeuLeuSerAspAspPheL 85
         !::::::|||||||||||||||||||          |||||||::::
     272 TAGTCTTTCATTGGACCAAAGTGAG.............. GATGACCTGT 306

      85 euAspThrProAspAspLeuAspIleAsnValAspAspIleGluThrPro 101
         ::|||||||||||||||||||||||||||||||:|| ||||:::::|||||
     307 TTGACACGCCAGATGACCTCGACATCAATGTGGATGATCTGGACACACCT 356

     102 AspGluThrAspSerLeuGluPheLeuGlyAsnGlyAsnGluLeuGluTr 118
         |||||||:::'||      :::               :::|||:::::||
     357 GATGAAGCAGATTATACAGAC.   .  ......CATGAAATAGACTG 391

     118 pGluAspAspThrProValAlaThrAlaLysAsnMetProGlyAspSerA 135
         :|:|:::      |||   .||:::   :::           :::::::
     392 GGAAGAA......CCTCAAGCATCCCAGCGAGTTTCTGTAAAAGAAACAG 435

     135 laAspLeuPheGlyAspGlyThrThrGluAspGlySerAlaAlaAsnGly 151
         :::::   :::         :::::::|||:::   :::      ::::::
     436 TTGAGCCCATTCCCACATACAGTGCTGAGGAG.. GAACGGCAGGATTCC 482

     152 ArgLeuTrpArgThrValIleIleGlyGluGlnGluHisArgIleAspLe 168
         :::|||||||||||||||||||||||||||:|||||:'||||||||||||.....
     483 AAACTCTGGAGAACGGTCATCATTGGAGAGCAAGAGCACCGCATCAACAT 532

     168 uHisMetIleArgProTyrMetLysValValThrHisGlyGlyTyrTyrG 185
         :    |||:::|||||||:::|||||||:::::|||||||||||||||||:
     533 GAAGAGCATTGAACCCTACCAAAAAGTCATATCCCATGGAGGCTACTATA 582

     185 lyGluGlyLeuAsnAlaIle 191
         :::::|||     ||||||
     583 GCAATGGTGCANATGCCATC 602
```